PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, 15/79, C07K 14/47, A61K 31/70, 48/00 — A1

(11) International Publication Number: WO 97/22696

(43) International Publication Date: 26 June 1997 (26.06.97)

(21) International Application Number: PCT/GB96/03156

(22) International Filing Date: 19 December 1996 (19.12.96)

(30) Priority Data:

9525962.8	19 December 1995 (19.12.95)	GB
9615797.9	26 July 1996 (26.07.96)	GB
9622174.2	24 October 1996 (24.10.96)	GB

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

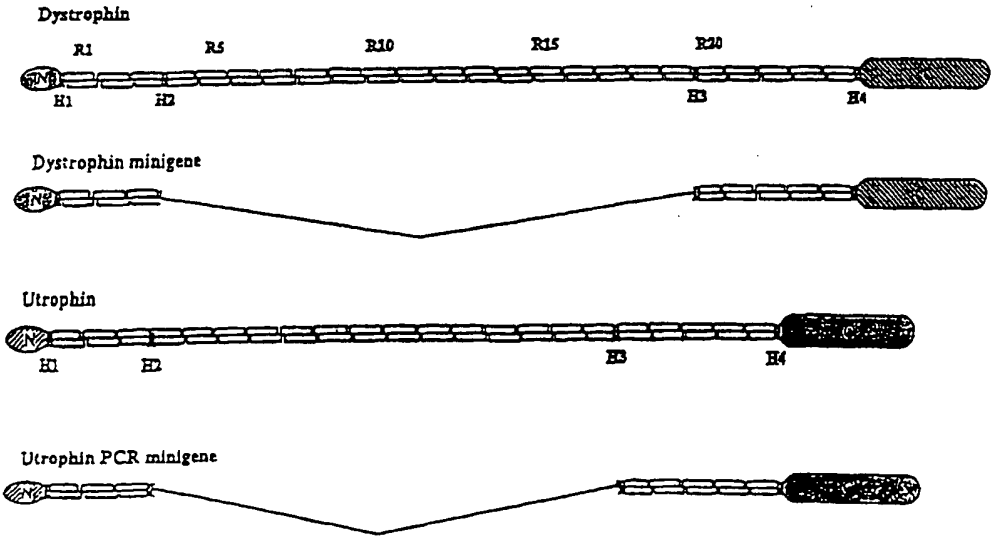
(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: TINSLEY, Jonathon, Mark [GB/GB]; Genetics Laboratory, Dept. of Biochemistry, South Parks Road, Oxford OX1 3QU (GB). DAVIES, Kay, Elizabeth [GB/GB]; Genetics Laboratory, Dept. of Biochemistry, South Parks Road, Oxford OX1 3QU (GB).

(74) Agents: WALTON, Seán, M. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

(54) Title: UTROPHIN GENE EXPRESSION



(57) Abstract

Nucleic acid from which a polypeptide with utrophin function can be expressed, especially mini-genes and chimaeric constructs. Expression significantly decreases the severity of the dystrophic muscle phenotype in an animal model, indicating usefulness in treatment of muscular dystrophy. The nucleic acid and encoded polypeptides are also useful in screening for substances to modulate utrophin binding to actin and/or the dystrophin protein complex.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

UTROPHIN GENE EXPRESSION

The present invention generally relates to the provision of nucleic acid from which a polypeptide with utrophin function can be expressed, especially mini-genes and chimaeric constructs. Expression of a utrophin transgene significantly decreases the severity of the dystrophic muscle phenotype in an animal model.

The severe muscle wasting disorders, Duchenne muscular dystrophy (DMD) and the less debilitating Becker muscular dystrophy (BMD) are due to mutations in the dystrophin gene. Dystrophin is a large cytoskeletal protein which in muscle is located at the cytoplasmic surface of the sarcolemma, the neuromuscular junction (NMJ) and myotendinous junction (MTJ). The protein is composed of four domains: an actin-binding domain (shown *in vitro* to bind actin), a rod domain containing triple helical repeats, a cysteine rich (CR) domain and a carboxy-terminal (CT) domain. The majority of the CRCT binds to a complex of proteins and glycoproteins (called the dystrophin protein complex, DPC) spanning the sarcolemma. This complex consists of cytoskeletal syntrophins and dystrobrevin, transmembrane, β-dystroglycan, α-,β-δ-,γ-sarcoglycans and extracellular α-dystroglycan. The DPC links to laminin-α2 (merosin) in the extracellular matrix and to the actin cytoskeleton *via* dystrophin within the cell. The breakdown of the integrity of the DPC due to the loss of, or impairment of dystrophin

function, leads to muscle degeneration and the DMD phenotype. The structure of dystrophin and protein interactions within the DPC have been recently reviewed [1,2,3].

There are various approaches which can be adopted for the gene therapy of DMD. These include myoblast transfer, retroviral infection, adenoviral infection and direct injection of plasmid DNA. In most cases the dystrophin gene used in the experiments generates a truncated protein approximately half the size of the full size protein. This dystrophin minigene was modelled on a natural mutation identified in a very mild Becker patient [4]. The cloned version of this truncated minigene is able to reverse the pathological phenotype in the dystrophin deficient *mdx* mouse [5,6,7] and has had limited success when delivered to *mdx* muscle by viral vectors [8,9,10]. Although some progress is being made in each of these areas using the *mdx* mouse as a model system, there are problems related to the number of muscle cells that can be made dystrophin positive, the levels of expression of the gene and the duration of expression [11]. Another problem to be addressed is the rejection of cells expressing dystrophin because of immunological intolerance i.e. dystrophin within these cells will appear foreign to the host immune system given that most DMD patients will never have expressed dystrophin [12,13].

In order to circumvent some of these problems, possibilities of compensating for dystrophin loss using a related protein, utrophin, are being explored.

Utrophin is a 395kDa protein encoded by a gene located on chromosome 6q24 and shown to have strong sequence similarity to dystrophin [14]. The actin binding domain of dystrophin and utrophin has 85% similarity and the DPC binding region has 88% similarity. Both of these domains have been shown to function as predicted *in vitro*. The structure and potential protein interactions are described in detail in reviews [1,2,3].

There is a substantial body of evidence demonstrating that utrophin is capable of localising to the sarcolemma. During normal fetal muscle development there is increased utrophin expression, localised to the sarcolemma up until 18 weeks and 20 days gestation in human and mouse respectively. After this time the utrophin sarcolemmal staining steadily decreases to the significantly lower adult levels shortly before birth where utrophin is localised almost exclusively to the NMJ and MTJ [15,16,17]. The decrease in utrophin expression coincides with increased expression of dystrophin [17]. Many studies have shown that utrophin is bound to the sarcolemma in DMD and BMD patients. However the levels of utrophin localised at the sarcolemma vary from report to report [18,19,20,21]. In some other non Xp21 myopathies, utrophin and

dystrophin are simultaneously bound to the sarcolemma of adult skeletal muscle [22].

High levels of utrophin may protect muscle from the consequences of dystrophin loss. Matsumara *et al.* [23] demonstrated that purified membranes from the *mdx* mouse contained complexes of utrophin and the DPC. When quadricep muscles (which show necrosis) from these mice were analysed by immunoblotting, the level of utrophin remained approximately the same, however the level of the α-dystroglycan was drastically reduced. In cardiac muscle (which shows no pathology) the level of utrophin was elevated four fold with no loss of the α-dystroglycan. Immunocytochemical analysis of other *mdx* small calibre skeletal muscles (extraocular and toe) which also have no pathology shows increased utrophin expression and normal levels of α-sarcoglycan. This result suggests that the increased levels of utrophin interacts with the DPC (or an antigenically related complex) at the sarcolemma and prevents loss of the complex thus the structure of these cells remains normal. In the *mdx* mouse, utrophin levels in muscle remain elevated soon after birth compared with normal mice; however once the utrophin levels have decreased to the adult levels (about 1 week after birth), the first signs of muscle fibre necrosis are detected [15,16].

Thus, in certain circumstances utrophin can localise to the sarcolemma probably at the same binding

sites as dystrophin, namely actin and the DPC. If the expression of utrophin is high enough, it may maintain the DPC and thus alleviate the DMD phenotype. It is unlikely that such external upregulation could be tightly controlled giving rise to potentially high levels of utrophin within the cell. However, this may not be a problem as Cox *et al.* [24] have demonstrated that gross over expression of dystrophin in the muscle of transgenic *mdx* mice reverts the muscle pathology to normal with no obvious detrimental side effects.

The present invention has arisen from cloning of nucleic acid encoding utrophin and fragments of utrophin from various species. The original aim was to clone nucleic acid encoding human utrophin, but major problems were encountered. A previous paper (14) reported the amino acid sequence of utrophin (so-called "dystrophin-related protein"), obtained by cloning of overlapping cDNAs. However, two regions around the amino terminal actin binding domain were not represented in these clones. These regions could be amplified by PCR and sequenced, but it has proved not to be possible to clone them. Either clones which should have included these regions were rearranged (as determined by restriction mapping) or simply no clones were isolated even if highly recombination deficient *E. coli* host strains (SURE and STBL2) were used. The gaps in the sequence were identified by comparing the sequence generated from the utrophin cDNAs to the

published human dystrophin sequence. It became apparent as further utrophin clones were isolated, none spanned these two gaps.

Sequence information obtained from the amino terminus of the human cDNA was used to design probes and rat and mouse cDNA libraries were screened. Rat cDNAs were also unstable or rearranged in the region corresponding to the unclonable regions in the human sequence. Some large rat clones covering these regions were obtained, but all attempts to generate subclones failed due to rearrangements of the inserts as determined by restriction mapping. Surprisingly, in view of the difficulties with the human and rat sequences, cDNA from the mouse library, covering the regions in question, was found to be stable and amenable to further manipulation including the generation of smaller subclones.

Figure 1 shows a comparison between human, rat and mouse utrophin nucleotide sequences encoding part of the amino-terminal portion of the respective proteins. The unclonable regions of the human gene are underlined.

This cloning work enables for the first time the construction of a nucleic acid molecule from which a polypeptide with utrophin function can be expressed.

Furthermore, by way of analogy with the success achieved with a dystrophin mini-gene (from which a truncated version of dystrophin is expressed) the

present invention provides "utrophin mini-genes" and polypeptides encoded thereby. To overcome the problem of unclonability of regions of the human utrophin gene sequence, the present inventors have realised that it is possible to employ a sequence of nucleotides derived from the mouse utrophin gene in a chimaeric construct to provide for expression of a polypeptide with utrophin function.

According to a first aspect of the present invention there is provided a nucleic acid molecule comprising a sequence of nucleotides encoding a polypeptide with utrophin function.

A polypeptide with utrophin function is able to bind actin and able to bind the dystrophin protein complex (DPC).

Polypeptides with utrophin function are generally distinguishable immunologically from dystrophin polypeptides. For example, they may comprise at least one epitope not found in dystrophin. Polypeptides with utrophin function may be identified using specific polyclonal or monoclonal antibodies which do not cross-react with dystrophin. If a polypeptide is able to bind actin and able to bind the dystrophin protein complex and at least one antibody can bind it which cannot bind dystrophin, then the polypeptide has utrophin function. In a preferred embodiment, the polypeptide can be bound by an antibody which binds utrophin but not dystrophin, in other words the

polypeptide shares at least one epitope with utrophin which epitope is not found in dystrophin. In another embodiment, the polypeptide does not contain an epitope found in dystrophin, such that the polypeptide is not bound by an antibody which binds dystrophin. In such a case, the epitope recognised by the antibody which binds dystrophin may be one not found in utrophin. The polypeptide may contain no epitope found in dystrophin. The immunological comparison may be made with human utrophin and/or dystrophin, especially if the polypeptide with utrophin function is intended for human use, or with the utrophin and/or dystrophin of the species in which use is intended, e.g. mouse. Mouse monoclonal antibodies MANCH07 and MANNUT1[31] were used in the work described herein. Standard *in vitro* binding assays may be used to assess immunological cross-reactivity of a polypeptide.

Thus, the polypeptide comprises an actin-binding domain and a dystrophin protein complex (DPC)-binding domain and utrophin-like as opposed to dystrophin-like, e.g. as determined immunologically.

Preferably the encoding sequence comprises a human sequence, i.e. a sequence obtainable from the genome of a human cell.

Comparison of various amino acid sequences reveals the following % similarities (calculated using the method of Needleman and Wunsch (1974) *J. Mol. Biol.* 48: 443-453, performed using the GAP program from

the Winsconsin Package v8, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin 53711, USA) and identities:

- full length human dystrophin v. human utrophin
  69% similarity, 50.7% identity;
- full length human utrophin v. rat utrophin
  93.2% similarity, 87.1% identity;
- full length human dystrophin v. mouse dystrophin
  95.4% similarity, 91.2% identity;
- human dystrophin C-terminus v. human utrophin C-terminus 84.1% similarity, 73.6% identity.

As noted, the present invention is only concerned with "utrophin-like" molecules, not "dystrophin-like" molecules. Thus, polypeptides according to the present invention (e.g. as encoded by nucleic acid according to the invention) may have an amino acid sequence which is greater than about 75% similar, preferably greater than about 80%, about 85%, about 90%, about 95% or about 98% similarity to the amino acid sequence of Figure 3 or the amino acid sequence of Figure 9, taken over the full length. The polypeptides may have an amino acid identity of greater than about 55% identity, preferably greater than about 60% identity, about 70%, about 80%, about 90%, about 95% or about 98% identity over the full-length. The levels of similarity and/or identity may be lower outside the C-terminal, DPC-binding domain provided the DPC-binding domain has greater than about 85% similarity, preferably greater than about 90%,

about 95% or about 98% similarity with the DPC-binding domain amino acid sequence of Figure 3 or Figure 9, or has greater than about 80%, preferably greater than about 85%, about 90%, about 95% or about 98% identity with the DPC-binding domain amino acid sequence of Figure 3 or Figure 9. Particular amino acid sequence variants or derivatives may have a sequence which differs from the sequence of Figure 3 or Figure 9 by one or more of insertion, addition, substitution or deletion of 1 amino acid, 2, 3, 4, 5-10, 10-20 20-30, 30-50, 50-100, 100-150, or more than 150 amino acids.

The nucleic acid molecule may be an isolate, or in an isolated and/or purified form, that is to say not in an environment in which it is found in nature, removed from its natural environment. It may be free from other nucleic acid obtainable from the same species, e.g. encoding another polypeptide.

The nucleic acid molecule may be one which is not found in nature. For example, the sequence of nucleotides may form part of a cloning vector and/or an expression vector, as discussed further below. The sequence of nucleotides may represent a variant or derivative of a naturally occurring sequence by virtue of comprising an addition, insertion, deletion and/or substitution of one or more nucleotides with respect to the natural sequence, provided preferably that the encoded polypeptide has the specified characteristics. The addition, insertion, deletion and/or substitution

of one or more nucleotides may or may not be reflected in an alteration in the encoded amino acid sequence, depending on the genetic code.

Preferably, the nucleic acid molecule is a "mini-gene", i.e. the polypeptide encoded does not correspond to full-length utrophin but is rather shorter, a truncated version. For instance, part or all of the rod domain may be missing, such that the polypeptide comprises an actin-binding domain and a DPC-binding domain but is shorter than naturally occurring utrophin. In a full-length utrophin gene, the actin-binding domain is encoded by nucleotides 1-739, while the DPC-binding domain (CRCT) is encoded by nucleotides 8499-10301 (where 1 represents the start of translation; Figure 2A). The respective domains in the polypeptide encoded by a mini-gene according to the invention may comprise amino acids corresponding to those encoded by these nucleotides in the full-length coding sequence.

Dystrophin mini-genes have been shown to be active in animal models (as discussed). Advantages of a mini-gene over a sequence encoding a full-length utrophin molecule or derivative thereof include easier manipulation and inclusion in vectors, such as adenoviral and retroviral vectors for delivery and expression.

A further preferred non-naturally occurring molecule encoding a polypeptide with the specified

characteristics is a chimaeric construct wherein the encoding sequence comprises a sequence obtainable from one mammal, preferably human ("a human sequence"), and a sequence obtainable from another mammal, preferably mouse ("a mouse sequence"). Such a chimaeric construct may of course comprise the addition, insertion, substitution and/or deletion of one or more nucleotides with respect to the parent mammalian sequences from which it is derived. Preferably, the part of the coding sequence which encodes the actin-binding domain comprises a sequence of nucleotides obtainable from the mouse, or other non-human mammal, or a sequence of nucleotides derived from a sequence obtainable from the mouse, or other non-human mammal.

In a preferred embodiment, the sequence of nucleotides encoding the polypeptide comprises sequence GAGGCAC at residues 331-337 and/or the sequence GATTGTGGATGAAAACAGTGGG at residues 1453-1475 (using the conventional numbering from the initiation codon ATG), and a sequence obtainable from a human.

The nucleic acid molecule may comprise a nucleotide sequence encoding a sequence of amino acids shown in Figure 1. As discussed, the encoding sequence may be chimaeric, i.e. comprise sequences of nucleotides from different species, e.g. a sequence from or derivable from a human and a sequence from or derivable from a mouse or other non-human mammal.

A chimaeric mini-gene encoding sequence according

to the present invention is shown in Figure 3. Preferred embodiments of the present invention include a nucleic acid molecule comprising a sequence of nucleotides encoding a polypeptide which has an actin-binding domain and a DPC-binding domain and which polypeptide comprises an amino acid sequence encoded by a sequence of nucleotides shown in Figure 3, a nucleic acid molecule comprising a sequence of nucleotides encoding a variant, allele or derivative of such a polypeptide by way of addition, substitution, insertion and/or deletion of one or more amino acids, and a nucleic acid molecule comprising a sequence of nucleotides which is a variant, allele or derivative of the sequence shown in Figure 3, by way of addition, substitution, insertion and/or deletion of one or more nucleotides, with or without a change in the encoded amino acid sequence with respect to the amino acid sequence encoded by a sequence of nucleotides shown in Figure 3. The proviso is that the encoded polypeptide is "utrophin-like" rather than "dystrophin-like", e.g. as determined immunologically as discussed.

One particular variant or derivative of the sequence of Figure 3 has a sequence as shown in Figure 9, which is a "full-length" utrophin construct, including rod domain sequences not included in the mini-gene of Figure 3.

The sequences of Figure 3 and Figure 9 include some positions at which the precise residue is left

open (marked by "N" in the nucleotide sequence and "X" in the amino acid sequence). Comparison of the human, mouse and rat utrophin sequences in this region (Figure 10) shows that the human and rat amino acid sequences are absolutely conserved here. Accordingly, the twelve "X's" in Figures 3 and 9 may represent the amino acid sequence DKKSIIMYLTSL. Instead, in accordance with the discussion of variants and derivatives herein, a polypeptide according to the invention (as encoded by nucleic acid according to the invention) may include a variant or derivative sequence, by way of one or more of insertion, addition, substitution or deletion of one or more amino acids of the sequence DKKSIIMYLTSL, in the position marked by the X's in Figures 3 and 9.

Nucleic acid according to the present invention is obtainable by hybridising nucleic acid of target cells (e.g. human, mouse, rat) with one or more oligo- or poly-nucleotides with sequences designed based on the sequence information presented in Figure 1, Figure 3 or Figure 9. Thus, the full mouse sequence, or the sequence in the region marked by the X's in Figures 3 and 9, may be obtained by probing or PCR using sequence information provided herein (e.g. Figure 1).

Nucleic acid according to the present invention is obtainable using one or more oligonucleotide probes or primers designed to hybridise with one or more fragments of a nucleic acid sequence shown in Figure 1, Figure 3 or Figure 9, particularly fragments of

relatively rare sequence, based on codon usage or statistical analysis. The amino acid sequence information provided may be used in design of degenerate probes/primers or "long" probes. A primer designed to hybridise with a fragment of the nucleic acid sequence shown may be used in conjunction with one or more oligonucleotides designed to hybridise to a sequence in a cloning vector within which target nucleic acid has been cloned, or in so-called "RACE" (rapid amplification of cDNA ends) in which cDNA's in a library are ligated to an oligonucleotide linker and PCR is performed using a primer which hybridises with the sequence shown in the figure and a primer which hybridises to the oligonucleotide linker.

Nucleic acid isolated and/or purified from one or more cells (e.g. human, mouse) or a nucleic acid library derived from nucleic acid isolated and/or purified from cells (e.g. a cDNA library derived from mRNA isolated from the cells), may be probed under conditions for selective hybridisation and/or subjected to a specific nucleic acid amplification reaction such as the polymerase chain reaction (PCR).

A method may include hybridisation of one or more (e.g. two) probes or primers to target nucleic acid. Where the nucleic acid is double-stranded DNA, hybridisation will generally be preceded by denaturation to produce single-stranded DNA. The hybridisation may be as part of a PCR procedure, or as

part of a probing procedure not involving PCR. An example procedure would be a combination of PCR and low stringency hybridisation. A screening procedure, chosen from the many available to those skilled in the art, is used to identify successful hybridisation events and isolated hybridised nucleic acid.

Probing may employ the standard Southern blotting technique. For instance DNA may be extracted from cells and digested with different restriction enzymes. Restriction fragments may then be separated by electrophoresis on an agarose gel, before denaturation and transfer to a nitrocellulose filter. Labelled probe may be hybridised to the DNA fragments on the filter and binding determined. DNA for probing may be prepared from RNA preparations from cells.

Preliminary experiments may be performed by hybridising under low stringency conditions various probes to Southern blots of DNA digested with restriction enzymes. Suitable conditions would be achieved when a large number of hybridising fragments were obtained while the background hybridisation was low. Using these conditions nucleic acid libraries, e.g. cDNA libraries representative of expressed sequences, may be searched.

It may be necessary for one or more gene fragments to be ligated to generate a full-length coding sequence. Also, where a full-length encoding nucleic acid molecule has not been obtained, a smaller

molecule representing part of the full molecule, may be used to obtain full-length clones. Inserts may be prepared from partial cDNA clones and used to screen cDNA libraries.

Those skilled in the art are well able to employ suitable conditions of the desired stringency for selective hybridisation, taking into account factors such as oligonucleotide length and base composition, temperature and so on.

Systems for cloning and expression of a polypeptide in a variety of different host cells are well known. Suitable host cells include bacteria, mammalian cells, yeast and baculovirus systems. Mammalian cell lines available in the art for expression of a heterologous polypeptide include Chinese hamster ovary cells, HeLa cells, baby hamster kidney cells and many others. A common, preferred bacterial host is *E. coli*.

Nucleic acid according to the present invention may form part of a cloning vector and/or a vector from which the encoded polypeptide may be expressed. Suitable vectors can be chosen or constructed, containing appropriate and appropriately positioned regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. Vectors may be plasmids, viral e.g. 'phage, or phagemid, as appropriate. For further

details see, for example, *Molecular Cloning: a Laboratory Manual*: 2nd edition, Sambrook et al., 1989, Cold Spring Harbor Laboratory Press. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Short Protocols in Molecular Biology*, Second Edition, Ausubel et al. eds., John Wiley & Sons, 1992. The disclosures of Sambrook et al. and Ausubel et al. are incorporated herein by reference.

Thus, a further aspect of the present invention provides a host cell containing nucleic acid as disclosed herein. A still further aspect provides a method comprising introducing such nucleic acid into a host cell. The introduction may employ any available technique. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation and transfection using bacteriophage.

The introduction may be followed by causing or allowing expression from the nucleic acid, e.g. by culturing host cells under conditions for expression of

the gene.

In one embodiment, the nucleic acid of the invention is integrated into the genome (e.g. chromosome) of the host cell. Integration may be promoted by inclusion of sequences which promote recombination with the genome, in accordance with standard techniques.

The invention also provides a mammal, such as a human, primate or rodent, preferably rat or mouse, comprising a host cell as provided, and methods of production and use of such a mammal. The mammal may be non-human. Transgenic animals, particularly mice, can be generated using any available technique. Particularly suitable for purposes of study are *mdx* mice or others with a dystrophic phenotype.

The polypeptide encoded by the nucleic acid may be expressed from the nucleic acid *in vitro*, e.g. in a cell-free system or in cultured cells, or *in vivo*. *In vitro* expression may be useful in determining ability of the polypeptide to bind to actin and/or DPC. This may be useful in testing or screening for substances able to modulate one or both of these binding activities. In particular, substances able to increase actin and/or DPC binding of the polypeptide will add to the repertoire of molecules available for potential pharmaceutical/therapeutic exploitation. Such substances, identified as modulators of one or both of the binding activities of the polypeptide, following

expression of the polypeptide from encoding nucleic acid therefor, may be investigated further and may be manufactured and/or used in preparation of a medicament, pharmaceutical composition or drug which may subsequently be administered to an individual. *In vivo* expression is discussed further below.

According to a further aspect of the present invention there is provided a polypeptide with utrophin function (other than utrophin itself). Such a polypeptide comprises an actin-binding domain and a DPC-binding domain and is immunologically recognisable as utrophin-like rather than dystrophin-like, as discussed, not-being a naturally occurring polypeptide. The polypeptide may be any of those discussed above as being encoded by nucleic acid according to the present invention. In particular, the polypeptide may be shorter than naturally occurring full-length utrophin, for example by virtue of lacking all or part of the rod domain. The actin-binding and DPC-binding domains may correspond to those of human, mouse or other non-human utrophin or may be derived therefrom by way of addition, substitution, insertion and/or deletion of one or more amino acids. The polypeptide may be chimaeric, comprising sequences of amino acids from or derived from different species, e.g. human and mouse, as discussed.

A convenient way of producing a polypeptide according to the present invention is to express

nucleic acid encoding it. Accordingly, methods of making such polypeptides by expression from encoding nucleic acid therefor are provided by the present invention, *in vitro*, e.g. in cell-free systems or by culturing host cells under appropriate conditions for expression, or *in vivo*.

Polypeptides and nucleic acid according to the invention may be used in the manufacture of medicaments, compositions, including pharmaceutical formulations, and drugs for delivery to an individual, e.g. a human with muscular dystrophy or a non-human mammal, such as a mouse, as a model for study of the polypeptides, muscular dystrophy and therapy thereof.

For example, a method of treatment practised on the human or animal body in accordance with the present invention may comprise administration to an individual of nucleic acid encoding a polypeptide as disclosed herein. The nucleic acid may form part of a construct enabling expression within cells of the individual. Nucleic acid may be introduced into cells using a retroviral vector, preferably one which will not transform cells, or using liposome technology.

Administration is preferably in a "therapeutically effective amount", this being sufficient to show benefit to a patient. Such benefit may be at least amelioration of at least one symptom. The actual amount administered, and rate and time-course of administration, will depend on the nature and

severity of what is being treated. Prescription of treatment, eg decisions on dosage etc, is within the responsibility of general practitioners and other medical doctors.

A composition may be administered alone or in combination with other treatments, either simultaneously or sequentially dependent upon the condition to be treated.

Pharmaceutical compositions according to the present invention, and for use in accordance with the present invention, may comprise, in addition to active ingredient, a pharmaceutically acceptable excipient, carrier, buffer, stabiliser or other materials well known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material will depend on the route of administration, which may be oral, or by injection, e.g. cutaneous, subcutaneous or intravenous.

Pharmaceutical compositions for oral administration may be in tablet, capsule, powder or liquid form. A tablet may comprise a solid carrier such as gelatin or an adjuvant. Liquid pharmaceutical compositions generally comprise a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene

glycol may be included.

For intravenous, cutaneous or subcutaneous injection, or injection at the site of affliction, the active ingredient may be in the form of a parenterally acceptable aqueous solution which is pyrogen-free and has suitable pH, isotonicity and stability. Those of relevant skill in the art are well able to prepare suitable solutions using, for example, isotonic vehicles such as Sodium Chloride Injection, Ringer's Injection, Lactated Ringer's Injection. Preservatives, stabilisers, buffers, antioxidants and/or other additives may be included, as required.

Injection may be used to deliver nucleic acid to disease sites. Internally, suitable imaging devices may be employed to guide an injecting needle to the desired site.

It may be desirable to remove cells from the body, treat them then return them to the body, or to administer cells derived from cells removed from an individual. This might be appropriate, for example, if muscle stem cells can be isolated. Muscle precursor cells ("mpc") have been used in cell therapy in *mdx* mice, where implantation of normal mpc gave rise to substantial amounts of dystrophin [25,26,27]. Immunosuppression increases success of cell implantation procedures [13]. Myoblasts may be used to introduce genes into muscle fibres during growth or repair, as has been demonstrated using a replication-

defective retroviral vector to introduce a mini-dystrophin construct into proliferating myogenic cells in tissue culture [28].

Thus, cells in culture may have nucleic acid according to the present invention introduced into them before the cells are grafted into muscles in a patient. Grafting the cells back into the donor has the advantages of a genetically corrected autologous transplant. Nucleic acid may be introduced locally into cells using transfection, electroporation, microinjection, liposomes, lipofecting or as naked DNA or RNA, or using any other suitable technique.

Retroviral vectors have also been used to introduce the dystrophin mini-gene into the myoblasts of spontaneously regenerating muscle of the *mdx* mouse to produce dystrophin-positive fibres [8]. Recombinant replication defective adenoviruses appear particularly effective as an efficient means of introducing constructs into skeletal muscle fibres for persistent expression [29]. See reference 11 for a review of myoblast-based gene therapies.

Adenoviral, retroviral or other viral vectors may be used advantageously for the introduction of a utrophin sequence according to the present invention into muscle cells. Even though *in vivo* transduction may be restricted to growing or regenerating muscle fibres, retrovirally introduced constructs have the advantage of becoming integrated into the genome of the

host cell, potentially conferring lifelong expression.

Liposomes may be used as vehicles for delivery of nucleic acid constructs to skeletal muscle. Intravenous injection of constructs in cationic liposomes has resulted in widespread transfection of most tissues, including skeletal muscle [30]. Lack of immunogenicity allows for repeated administration and lack of tissue specificity may be accommodated by choosing a muscle-specific promoter to drive expression.

For use in distinguishing polypeptide with utrophin function from dystrophin and related polypeptides, antibodies may be obtained using techniques which are standard in the art. Methods of producing antibodies include immunising a mammal (eg mouse, rat, rabbit, horse, goat, sheep or monkey) with the protein or a fragment thereof, or a cell or virus which expresses the protein or fragment. Immunisation with DNA encoding a target polypeptide is also possible (see for example Wolff, et al. *Science* 247: 1465-1468 (1990); Tang, et al. *Nature* 356: 152-154 (1992); Ulmer J B, et al. *Science* 259: 1745-1749 (1993)). Antibodies may be obtained from immunised animals using any of a variety of techniques known in the art, and screened, preferably using binding of antibody to antigen of interest. For instance, Western blotting techniques or immunoprecipitation may be used (Armitage et al, 1992, Nature 357: 80-82).

The production of monoclonal antibodies is well established in the art. Monoclonal antibodies can be subjected to the techniques of recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementarity determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. See, for instance, EP184187A, GB 2188638A or EP-A-0239400. A hybridoma producing a monoclonal antibody may be subject to genetic mutation or other changes, which may or may not alter the binding specificity of antibodies produced.

As an alternative or supplement to immunising a mammal with a peptide, an antibody specific for a protein may be obtained from a recombinantly produced library of expressed immunoglobulin variable domains, eg using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on their surfaces; for instance see WO92/01047. The library may be naive, that is constructed from sequences obtained from an organism which has not been immunised with the target, or may be one constructed using sequences obtained from an organism which has been exposed to the antigen of interest (or a fragment thereof).

Antibodies may be modified in a number of ways. Indeed the term "antibody" should be construed as covering any specific binding substance having an binding domain with the required specificity. Thus this covers antibody fragments, derivatives, functional equivalents and homologues of antibodies, including any polypeptide comprising an immunoglobulin binding domain, whether natural or synthetic. Chimaeric molecules comprising an immunoglobulin binding domain, or equivalent, fused to another polypeptide are therefore included. Cloning and expression of chimaeric antibodies are described in EP-A-0120694 and EP-A-0125023.

It has been shown that the function of binding antigens can be performed by fragments of a whole antibody. Example binding fragments are (i) the Fab fragment consisting of VL, VH, CL and CH1 domains; (ii) the Fd fragment consisting of the VH and CH1 domains; (iii) the Fv fragment consisting of the VL and VH domains of a single antibody; (iv) the dAb fragment (Ward, E.S. et al., Nature 341, 544-546 (1989)) which consists of a VH domain; (v) isolated CDR regions; (vi) F(ab')2 fragments, a bivalent fragment comprising two linked Fab fragments (vii) single chain Fv molecules (scFv), wherein a VH domain and a VL domain are linked by a peptide linker which allows the two domains to associate to form an antigen binding site (Bird et al, *Science*, 242, 423-426, 1988; Huston et al, *PNAS USA*,

85, 5879-5883, 1988); (viii) bispecific single chain Fv dimers (PCT/US92/09965) and (ix) "diabodies", multivalent or multispecific fragments constructed by gene fusion (WO94/13804; P. Holliger et al Proc. Natl. Acad. Sci. USA 90 6444-6448, 1993).

Further aspects and embodiments of the present invention, and modifications to aspects and embodiments disclosed herein, will be apparent to those skilled in the art.

The following figures are attached hereto:

Figure 1: Figure 1a shows the corresponding parts of nucleotide sequences of the mouse (Moutro), rat (ratutro) and human (humutro) starting from the first amino acid and encompassing the actin binding domain and start of the rod domain. The heavyline represents the unclonable region in rat and human. Figure 1b shows the sequence of the second unclonable region.

Figure 2 is a schematic of the cloning process of PCR6.0. Figure 2A: The bold numbers represent the utrophin transcript in kb and the numbers below the line represent the nucleotide positions of the regions in question where 1 is the start of translation; Figure 2B represents the cDNAs used as template for the PCR; Figure 2C represents the two PCR fragments generated to form PCR6.0.

Figure 3 shows the nucleotide sequence (both

strands) of a "utrophin mini-gene" according to the present invention and whose construction is described herein.

Figure 4 shows a representation of the dystrophin and utrophin polypeptides, showing the various domains, and "mini-genes" comprising only parts of the full-length molecules.

Figure 5: Utrophin transgene construction and expression. A, Scale representation of dystrophin, utrophin and the two truncated transgenes. The repeated spectrin-like repeats (R) and the potential hinge sites (H) are marked. B, Utrophin transgene vector. The N- and C-terminal portions of utrophin were cloned as PCR products using overlapping cDNAs as template. The regions used are indicated by the dotted lines. The PCR product was cloned into a vector containing the 2.2kb human skeletal α-actin (HSA) promoter and regulatory regions [20,21] and SV40 large T poly A site. The cloning sites were such that the transgene was located near the beginning of the second HSA untranslated exon and the Asp718/NotI sites were used to liberate the complete fragment. C, Immunoblot of muscle from the utrophin transgenic line F-3 and a non-transgenic C57BL/10 littermate. M, skeletal muscle; H, heart; D, diaphragm.

Figure 6: Decrease in serum CK levels and centralised myofibres in transgenic *mdx* mice. A, Serum creatine kinase levels in male *mdx* mice expressing the

utrophin transgene. Serum creatine kinase levels from 5 week old mice generated from 4 F_3 litters resultant from a male transgenic mouse crossed with female *mdx*. Offspring consisted of male hemizygous *mdx* (M *mdx*), male utrophin transgenic *mdx* (M Tg *mdx*) and heterozygous females (F *mdx*). Female heterozygotes are not significantly different from wild type so can be used as normal controls. The number of mice (n) in each group is shown in parentheses and the mean SE shown by T-bars. B, Proportion of myofibres containing centralised nuclei. The mean SE is shown by T-bars.

Figure 7: Decrease in centralised nuclei in Diaphragm and TA muscle from other truncated utrophin transgenic *mdx* lines (Gerald, George, Grant, Gavin), normal (n) and *mdx* (mdx). The mean SE is shown by T-bars.

Figure 8: Decrease in serum creatine kinase from other utrophin transgenic lines, normal (n) and *mdx* (mdx). The number of mice in each group is shown in parenthesis.

Figure 9: Full length utrophin coding sequence and encoded amino acid sequence.

Figure 10: Alignment of amino acid sequences for the N-terminal regions of human, mouse and rat utrophin.

All documents cited are incorporated herein by reference.

Example 1

Cloning of utrophin minigenes

The four cDNAs covering the latter half of the human utrophin transcript were ligated together using overlapping restriction endonuclease sites. The amino-terminal region was reconstructed using the human 92.2 cDNA joined by the common EcoRI restriction site to the stable mouse cDNA clone, JT1. These two constructs were then used as templates for PCR amplification (Figure 2B). Primers were designed to generate two fragments, PCR2.0 and PCR 4.0, containing no untranslated regions which could be ligated in frame to generate a utrophin minigene containing approximately the first 2kb and last 4kb of the utrophin coding sequence (Figure 2C).

The two PCR fragments were ligated together using the HpaI site. The complete DNA sequence of the 6.0kb minigene is shown in Figure 3. The complete 6kb minigene was excised from the vector and ligated into the eukaryotic expression vectors. SV40-pA consists of the SV40 early promoter linked to exon 1 and part of exon 2 (including the intron) of rabbit β globin to facilitate splicing of any cloned insert. This is of particular importance if the construct is to be used to generate transgenic lines. After a single unique blunt restriction site for cloning inserts into, there is the SV40 small T poly A signal sequence. The SV40 promoter

will express the minigene in all tissues. The HSA-pA construct is similar except for the use of the human skeletal α actin promoter and tissue specific regulatory sequences which will direct expression of the minigene product only in skeletal muscle.

Once cloned into the expression vectors the unique HpaI site was used to clone in a PCR generated fragment containing the remainder of the utrophin rod domain. We now have expression vectors containing a truncated and full length utrophin coding sequence.

The unique HpaI restriction site has also been used to clone in a synthetic oligonucleotide coding for the amino acid sequence which is recognised by a specific antibody to the myc protein. This will enable minigene constructs to be localised by virtue of their expression of the myc tag and recognition by the antibody. For utrophin this is a problem as the endogenous gene is expressed in all cell types. The use of the tag will demonstrate the presence of the minigene when delivered in a gene therapy protocol. There are available a number of other tags including the Flag epitope (IBI) and Green Fluorescent Protein (Clontech) which could be used in a similar fashion.

The utrophin minigene generated consists of the same domains and repeats as the dystrophin minigene (Figure 4). The dystrophin minigene was originally copied *in vitro* from a naturally occurring dystrophin mutation which gave rise to a mild Becker muscular

dystrophy phenotype. It has been used successfully in a number of viral vectors being designed for potential gene therapy routes and in transgenic lines which ameliorate the abnormal muscle phenotype in *mdx* mice. Thus the utrophin minigene will be suitable for cloning into viral vectors designed for specific tissue expression in potential gene therapy procedures.

Verification of the integrity of minigenes

It was important to screen for the maintenance of an open reading frame in the PCR generated clones given the propensity for *Taq* thermostable polymerase to introduce mutations. The PCR products were cloned into a vector which had RNA polymerase binding sites allowing the cloned insert to be transcribed and translated generating a radiolabelled protein. If expressed proteins were observed of the correct molecular weight it was inferred that the PCR product had no stop mutations. These products were then western blotted to see if they were recognised by utrophin antibodies. A positive result demonstrated that the expressed protein was in the correct frame to generate the epitopes recognised by the antibodies. Ten different clones both for the PCR2.0 and PCR4.0 were screened in this manner. In all cases full length expression was observed. All PCR2.0 and PCR4.0 clones were detected by MANNUT1 [31] (which recognises the actin binding domain) and MANCHO7 [18](which recognises

the latter half of the carboxy-terminal) respectively.

Two minigenes were constructed from two different PCR2.0 and 4.0 clones which met the criteria above and cloned into the expression vectors. To check the integrity of the completed minigene, COS cells were transiently transfected with both SV40-PCR6.0 minigenes (A4 and B1) and harvested after time points. Expression of the PCR6.0 minigene protein was identified by western blotting using MANNUT1 [31] and MANCH07 [18].

Similar transfections were done using these constructs then the cells fixed and immunostained using MANCHO7 [18]. Staining of the minigene appeared to be membrane bound suggesting that the actin binding domain or the CRCT or both are functional in order to explain the staining pattern seen.

The myc tag epitope has also been cloned in frame into the unique HpaI site within the minigene. This construct, SV40-PCR6.0-myc, was also transfected into COS cells and immunolocalised using the myc tag mouse monoclonal antibody, 9E10. Again membrane localisation was observed showing that introduction of the 10 amino acids which constitutes the myc tag epitope does not appear to effect the properties of the minigene.

Example 2

In Vivo Compensation for Dystrophic Deficiency by

Utrophin Expression

We have tested expressing a utrophin transgene in the dystrophin deficient *mdx* mouse. Our results indicate that high expression of the utrophin transgene in skeletal muscle can reverse the dystrophic pathology. These data suggest that systemic up-regulation of utrophin in DMD patients is a very promising avenue for the development of an effective treatment for this devastating disorder.

A truncated utrophin transgene was modeled on the Becker dystrophin transgene which has been shown to correct the dystrophic phenotype of *mdx* mice [5,6] (Fig. 5A). In order to generate high levels of muscle expression the utrophin transgene was driven by the human skeletal alpha actin (HSA) promoter (Fig. 5B). A number of transgenic lines expressing the utrophin transgene were generated with differing levels of transgenic expression. Immunoblot analysis of muscle samples from transgenic lines demonstrating high level expression are shown in Fig. 5C. The multiple fainter bands are probably due to the proteolytic breakdown of the highly expressed transgene product [24]. Line 347 also shows weak expression of the transgene in the heart. Analysis of the F-3 line shows no evidence of transgene expression in heart, brain, kidney, lung, liver, intestine, skin or pancreas was observed. To demonstrate that the utrophin transgene localised to the sarcolemma, immunofluorescence of skeletal muscle

sections was performed using utrophin and dystrophin specific antibodies. Examination of the sarcolemmal localisation pattern of dystrophin and the utrophin transgene in consecutive muscle sections demonstrated that they are able to co-localise *in vivo*. The normal localisation of utrophin in adult skeletal muscle is exclusively at the neuromuscular and myotendenous junctions and in the capillaries and nerves [3,31]. Immunostaining of unfixed 8μm TA muscle cryosections was done with 1/25 dilution of G3 (anti-utrophin) or 1/400 dilution of P6 (anti-dystrophin [33]). Initially the sections were blocked in 10% heat inactivated foetal calf serum in 50mM Tris, 150mM NaCl pH7.5 (TBS), then the primary antibody diluted in TBS added and incubated for 1h at room temperature. The slides were washed 4x in TBS for 5min each then incubated for a further hour at room temperature with 1/1000 dilution FITC conjugated sheep anti-rabbit IgG (Sigma) diluted in TBS. Finally the slides were washed as before, mounted with VectaShield (Vector Labs) and photographed using a Leica DMRBE microscope and photomicrograph system.

Although the dystrophin deficient *mdx* mouse is only mildly affected, histological and physiological analysis reveals a number of muscle defects in common with DMD patients including muscle fibre degeneration giving rise to a dramatic elevation of serum creatine kinase (CK) and evidence of massive myofibre

regeneration with most fibres having centrally located nuclei [34] Thus changes in the levels of serum CK and numbers of centralised nuclei have been used to monitor the pathology of the muscle in a number of transgenic lines expressing dystrophin transgenes in *mdx* mice [5,6,7,24]. Male transgenic F-3 mice carrying the utrophin transgene were crossed with dystrophin deficient female *mdx* mice and the resultant offspring analysed (Fig. 6A). The CK levels of 5 week old male transgenic *mdx* mice had fallen to approximately a quarter of the non transgenic *mdx* male littermates. Females whether transgenic or not have essentially normal levels of serum CK. The reduction in the serum levels of CK in the transgenic male *mdx* littermates signifies a change in the muscle pathology of these mice and implies that a significant decrease in muscle degeneration has occurred. Fig. 6B shows the contrast in numbers of centralised nuclei in frozen sections from the soleus and *tibialis anteria* (TA) muscle of transgenic and non-transgenic male *mdx* mice. The numbers of centrally nucleated myofibres is markedly reduced in the two muscle types examined showing that the amount of fibre regeneration is decreased. The difference in numbers of central nuclei between the transgenic *mdx* TA (~10%) and soleus (~30%) is probably explained by the fact that the HSA promoter is expressed at lower levels in the slow twitch fibres which essentially populate the soleus muscle compared

to the fast twitch fibres of the TA. This is an important observation as it implies that the levels of utrophin transgene are important for amelioration of the muscle phenotype.

Dystrophin is normally associated with a large oligomeric protein complex (dystrophin protein complex; DPC) embedded in the sarcolemma [3,35]. Loss of dystrophin in DMD patients and *mdx* mice also results in a dramatic loss of sarcolemmal DPC [36]. In transgenic *mdx* mice expressing the full length and truncated dystrophin transgenes, re-establishment of components of the DPC at the sarcolemma is an important marker for the restoration of muscle strength by dystrophin transgenes [5,6,7,24]. We looked at the results of immunostaining for components of the DPC in TA muscle from male *mdx* or *mdx* expressing the utrophin transgene. This was as described above. The primary antibodies were goat polyclonal sera to α/β-dystroglycan [37] (FP-B, 1/10), rabbit polyclonal sera to α-sarcoglycan [38] (1/5) and sheep polyclonal sera to γ-sarcoglycan [39] (1/10). FITC conjugated secondary antibodies to goat, rabbit and sheep were diluted 1/50, 1/200 and 1/50 respectively. Sarcolemmal staining of all myofibres by utrophin specific antibody was seen in transgenic muscle. However in the non-transgenic mice there is virtually no sarcolemmal staining apart from neuromuscular junctions and regions likely to contain regenerating fibres. In all cases using polyclonal

antibodies specific to α-sarcoglycan , γ-sarcoglycan and α/β-dystroglycan there was a notable increase in the staining at the sarcolemma of transgenic TA muscle indicating an elevation in correctly localised, sarcolemmal bound DPCs. The increase in sarcolemmal staining of these components in the soleus muscle is greater than the non-transgenic *mdx* males but not as elevated as in the TA. This result suggests that increased utrophin transgene expression correlates with an increase in sarcolemmal bound DPC.

Analysis of the *mdx* diaphragm has shown that this muscle exhibits a continued pattern of degeneration, fibrosis and functional deficit throughout the lifespan of the *mdx* mouse which is comparable to DMD skeletal muscle [40]. Thus for utrophin to be capable of replacing dystrophin, over-expression of utrophin in this muscle has to alter the pathology in a similar way as demonstrated for the dystrophin transgenic *mdx* mice [5,6,7,24]. Immunostaining of diaphragm sections using a utrophin antibody demonstrates the sarcolemmal localisation of the utrophin transgene expressed in the transgenic *mdx* mouse (utro-tg mdx) compared to the normal and *mdx* seen is the re-establishment of α-sarcoglycan at the sarcolemma of the transgenic *mdx* diaphragm at levels similar to the normal diaphragm. Sarcolemmal staining of the transgenic *mdx* diaphragm similar to normal is also seen using antibodies specific to α/β-dystroglycan and γ-sarcoglycan (data

not shown). Thus, as in skeletal muscle, expression of the utrophin transgene in diaphragm relocalises the DPC to the sarcolemma. Histological analysis of haematoxylin and eosin stained sections of *mdx* diaphragm shows extensive regions of fibrosis, cellular infiltration and variable myofibre size containing centralised nuclei. However the utrophin transgenic diaphragm looks essentially the same as normal, with no necrosis, regular myofibre size and virtually no centralised nuclei. In the *mdx* diaphragm, even in regions which have no necrosis so appear more histologically normal, on higher magnification the myofibres are still of variable size often containing centralised nuclei which is indicative of continual regeneration. In the utrophin transgenic diaphragm even at higher magnification, the whole muscle appears normal. A return to normal histology and establishment of the DPC are two important observations, as seen with the dystrophin transgenic mice [5,6,7,24], which predicts a major recovery of the utrophin transgenic diaphragm from a dystrophic phenotype.

We have demonstrated a significant decrease in the dystrophic muscle phenotype of *mdx* mice by expressing a utrophin transgene at high levels in the skeletal muscle and the diaphragm. These results, for the first time, strongly suggest that utrophin can replace dystrophin *in vivo* . This implies that use of small molecules which increase the normal utrophin

muscle expression to compensate and therefore alleviate the consequences of a lack of dystrophin is a promising avenue for DMD therapy. This approach would potentially target all muscles and thus prolong life by conserving the respiratory and cardiac muscles. Utrophin is expressed in many tissues so a generalised upregulation may not have detrimental side effects. In our experimental animal model, the normal mice expressing the utrophin transgene at high levels appear to suffer no deleterious side effects in their skeletal and diaphragm muscles. A precedent for such a gene therapy approach using butyrate to upregulate fetal haemoglobin is having success in clinical trials of sickle cell disease [41,42]. Only 20-30% of the wild type levels of dystrophin are required to significantly reduce the dystrophic phenotype in *mdx* mice [9,10]. It will be interesting to determine whether similar levels of utrophin will be adequate to compensate for dystrophin loss. In addition, since utrophin is normally expressed in all tissues, including muscle, the use of this utrophin transgene rather than a dystrophin transgene in conventional gene therapy approaches e.g. using viruses or liposomes may avert any potential immunological responses against the transgene.

Muscle from transgenic *mdx* and *mdx* mice were stressed *in vitro*. Essentially the test monitors the high mechanical stress produced by force lengthening

during active contraction and measures the decrease in force. The method is essentially described in detail by Deconinck *et al* [46]. The measure of deterioration is a decrease in the force a muscle can apply. This force drop is irreversible and correlates with the number of damaged muscle fibres. *Mdx* muscle is particularly sensitive to this test and deteriorates greatly [46].

Our data demonstrate that the force drop in *mdx* mice is -55%. However, in the utrophin transgenic littermates the force drop was only -20%. Normal mouse muscle usually has a force drop of -15%. Thus the expression of the utrophin transgene in *mdx* mice considerably decreases the damage caused by large mechanical stress.

Methods

Transgene construction and microinjection

The amino- and carboxy-terminal portions of utrophin were cloned as PCR products using overlapping cDNAs as template then ligated together in-frame to produce the truncated utrophin cDNA. The PCR product was then cloned into a vector containing the 2.2kb human skeletal α-actin (HSA) promoter and regulatory regions [43,44] and SV40 large T poly A site. The cloning sites were such that the transgene was located near the beginning of the second HSA untranslated exon

and the Asp718/NotI sites were used to liberate the complete fragment. Transgenic mice were generated by microinjection of the purified HSA transgene insert into the pronucleus of F_2 hybrid oocytes from C57BL/6xCBA/CA parents [45]. Positive transgenic mice were identified by southern blotting using a probe to the central part of the utrophin transgene. A number of founder F_0 males were bred to generate more offspring for analysis and breeding.

Protein analysis

Total muscle extracts were prepared by homogenisation in 1ml extraction buffer (75mM Tris pH6.8, 3.8% SDS, 4M Urea, 20% Glycerol, 5% β-mercaptoethanol) then heated 95°C for 5min. Usually 50μg of total protein (quantitated using Biorad DC protein assay kit) was loaded onto 6% polyacrylamide gels and transferred to nitrocellulose. Utrophin transgene expression was detected using a 1/200 dilution of mouse anti-utrophin monoclonal antibody (MANCHO7 [18]) and visualised using anti-mouse IgG-POD and chemiluminescence (Boehringer). For sectioning, skeletal muscle samples were removed and immersed in OCT compound (BDH) and frozen in liquid nitrogen cooled isopentane. Diaphragm was removed, cut in half then rolled longitudinally and sandwiched between Ox liver to facilitate orientation and easier sectioning. The sandwich was then frozen. Immunostaining of unfixed

8μm cryosections was performed by blocking the sections in 10% heat inactivated foetal calf serum in 50mM Tris, 150mM NaCl pH7.5 (TBS), then the primary antibody diluted in TBS added and incubated for 1h at room temperature. The slides were washed 4x in TBS for 5min each then incubated for a further hour at room temperature with conjugated second antibody diluted in TBS. Finally the slides were washed as before, mounted with VectaShield (Vector Labs) and photographed using a Leica DMRBE microscope and photomicrograph system.

Antibodies used for immunofluorescence

Antibodies were used at the following dilutions. Polyclonal rabbit against utrophin (G3, 1/25), dystrophin (P6 [33], 1/400), β1-syntrophin (syn35, 1/50) α-sarcoglycan [38] (1/5). Goat polyclonal against α/β-dystroglycan (FP-B [37], 1/10). FITC conjugated secondary antibody to goat (Sigma) and Cy3 conjugated secondary antibody to rabbit (Jackson Laboratories) were diluted 1/50 and 1/200 respectively.

Creatine kinase assay

Serum CK levels from 4-5 week old mice generated from 4 F_3 litters resultant from a male transgenic mouse crossed with female *mdx* were assayed. The tail tips were cut off and DNA prepared for Southern blotting to establish the transgenic status of each mouse. Blood was collected simultaneously, allowed to

clot and serum removed. Serum creatine kinase levels were measured using the Boehringer NAC-CK kit and 5μl of serum. The rate per minute was averaged over 4 min and calculated as U/l.

Example 3

To see if expression of utrophin is beneficial to muscle in the process of regenerating, the *myc* tagged truncated utrophin minigene under the control of the HSA promoter (HSA-PCR6.0-myc) was directly injected into *mdx* muscle.

Our data demonstrate the sarcolemmal localisation of the utrophin minigene in a proportion of fibres close to the injection site. The utrophin minigene was detected using the antibody 9E10 which is specific to the *myc* tag epitope. Importantly where 9E10 was localised there was a significant staining of α- and γ-sarcoglycan. The α-sarcoglycan staining was essentially negative in other fibres.

Re-establishment of the dystrophin protein complex has been shown to be an important marker for muscle recovery [5,6,7,24]. This result suggests that even when the disease process has manifested itself, namely the degeneration and regeneration seen in *mdx* muscle, expression of utrophin is beneficial. This is important when considering that in DMD one third of effected boys are new mutations. Thus only when the

first symptoms of DMD manifest themselves after a couple of years after birth can diagnosis be attained.

Example 4

Utilising a PCR strategy to generate fragments from human 1st strand DNA, the remainder of the human utrophin sequences missing from PCR6.0 were cloned. The fragment utilised primers which allowed the rod domain to be cloned into the unique HpaI restriction site (see Figure 2c) to produce a clone which contained all of the amino acid coding sequence to produce the complete utrophin protein. Figure 9 shows the DNA sequence of the full length utrophin construct with the amino acid sequence shown above using the standard single letter code.

The utrophin full length construct has been cloned into the human skeletal alpha actin promoter (HSA) expression construct in a similar manner to that shown in Figure 5b. This full length utrophin expression construct has been used to generate transgenic mice capable of expressing the full length utrophin protein in mouse muscle. Similar experiments may be performed as described in Example 2 to identify any differences in the effectiveness of the full length utrophin protein compared with the truncated utrophin protein in alleviating the muscle pathology in *mdx* mice.

In order to assess whether high levels of utrophin expression in all tissues is detrimental, to assist in planning therapeutic protocols and in particular choosing between tissue-specific expression or non-specific expression, a mouse model is being developed using the full length utrophin construct. Transgenic mice will be created expressing the full length utrophin protein under the regulation of a promoter which is expressed in all tissues. The promoter chosen for the first experiments is the human Ubiquitin-C promoter which has been shown to express in all tissues. Once these mice are shown to be expressing the full length utrophin transgene they will be monitored to identify any potential side effects caused by abnormally high levels of utrophin.

References

1. Winder *et al.* (1995) *FEBS Letts* 369: 27-33.
2. Blake *et al.* (1994) *Trend In Cell Biol* 4: 19-23.
3. Tinsley *et al.* (1994) *Proc Natl Acad Sci, USA* 91: 8307-8313.
4. England *et al.* (1990) *Nature* 11: 180-182.
5. Phelps *et al.* (1995) *Hum Mol Genet* 4: 1251-1258.
6. Wells *et al.* (1995) *Hum Mol Genet* 4: 1245-1250.
7. Rafael *et al.* (1994) *Hum Mol Genet* 3: 1725-1733.
8. Dunckley *et al.* (1993) *Hum Mol Genet* 2: 717-723.
9. Alameddine *et al.* (1994) *Neuromusc Dis* 4: 193-

203.

10. Vincent *et al.* (1993) *Nat Genet* 5: 130-134.

11. Partridge TA and Davies KE. (1995) *Brit Med Bull* 51: 123-137.

12. Guerette *et al.* (1995) *Muscle Nerve* 18: 39-51.

13. Huard *et al.* (1994) *Muscle Nerve* 17: 224-234.

14. Tinsley *et al.* (1992) *Nature* 360: 591-593.

15. Khurana *et al.* (1991) *Neuromusc Disord* 1: 185-194.

16. Takemitsu *et al.* (1991) *Biochem Biophys Res Com* 180: 1179-1186.

17. Clerk *et al.* (1993) *Histochem* 25: 554-561.

18. Nguyen *et al.* (1991) *J Cell Biol* 115: 1695-1700.

19. Tanaka *et al.* (1991) *Histochem* 96: 1-5.

20. Karpati *et al.* (1993) *J Neuropath Exp Neurology* 52: 119-128.

21. Mizuno *et al.* (1993) *J Neurol Sci* 1993.

22. Helliwell *et al.* (1992) *Neuromusc Dis* 3: 177-184.

23. Matsumura *et al.* (1992) *Nature* 360: 588-591.

24. Cox *et al.* (1993) *Nature* 364: 725-729.

25. Patridge *et al.* (1989) *Nature* 337: 176-179.

26. Karpati *et al.* (1989) In *Cellular and Molecular Biology of Muscle Development* (Alan R. Liss, New York) pp973-985).

27. Morgan *et al.* (1990) *J Cell Biol* 111: 2437-2449.

28. Dunckley *et al.* (1992) *FEBS Letts* 296: 128-134.

29. Stratford-Perricaudet *et al.* (1992) *J Clin Invest* 90: 626-630.

30. Thierry *et al.* (1995) *Proc Nat Acad Sci USA* 92: 9742-9746.

31. Ngygen *et al.* (1995) *FEBS Letts* 358: 262-266.

32. Blake, et al. *Brain Pathology* 6, 37-47 (1996).

33. Sherratt, et al. *Biochem J* 287, 755-759 (1992).

34. Sicinski, et al. *Science* 244, 1578-80 (1989).

35. Matsumara, K. & Campbell, K.P. *Muscle Nerve* 17, 2-15 (1994).

36. Campbell, K.P. *Cell* 80, 675-9 (1995).

37. Ibraghimov-Beskrovnaya, et al. *Nature* 355, 696-702 (1992).

38. Roberds, et al. *Biol. Chem.* 268, 23739-23742 (1993).

39. Jung, et al, *FEBS Letts.* 381, 15-20 (1996).

40. Stedman, et al. *Nature.* 352, 536-539 (1991).

41. Collins, et al. *Blood* 85, 43-9 (1995).

42. Dover, et al. *Blood* 84, 339-43 (1994).

43. Brennan, K.J. & Hardeman, E.C. *J. Biol. Chem.* 268, 719-725 (1993).

44. Muscat, G.O. & Kedes, L. *Mol. Cell Biol.* 7, 4089-4099 (1987).

45. Hogan, et al. *Manipulating the mouse embryo: A laboratory manual* (Cold Spring Harbour Laboratory Press, New York, 1986).

46. Deconinck, *et al. Proc. Natl. Acad. Sci. USA* 93, 3570-3574 (1996).

CLAIMS

1. A nucleic acid isolate having a sequence of nucleotides encoding a polypeptide including the amino acid sequence of Figure 3.

2. Nucleic acid according to claim 1 wherein said sequence of nucleotides is the coding nucleotide sequence of Figure 3.

3. Nucleic acid according to claim 1 wherein said sequence of nucleotides is a variant or derivative, by way of one or more of addition, substitution, insertion and deletion of one or more nucleotides, of the coding nucleotide sequence of Figure 3.

4. A nucleic acid isolate having a sequence of nucleotides encoding a polypeptide which is able to bind actin and able to bind the dystrophin protein complex, which includes an amino acid sequence which is a variant or derivative, by way of one or more of addition, substitution, insertion and deletion or one or more amino acids, of the amino acid sequence of Figure 3, and which is distinguishable immunologically from dystrophin.

5. Nucleic acid according to claim 4 wherein the polypeptide includes the amino acid sequence of Figure

9.

6. Nucleic acid according to claim 5 having the coding nucleotide sequence of Figure 9.

7. Nucleic acid according to claim 5 having a coding nucleotide sequence which is a variant or derivative, by one or more of addition, deletion, insertion or substitution of one or more nucleotides, of the coding sequence of Figure 9.

8. Nucleic acid according to any preceding claim comprised in a vector.

9. Nucleic acid according to claim 8 wherein said vector is an expression vector.

10. A composition including nucleic acid according to any of claims 1 to 9 and a pharmaceutically acceptable excipient.

11. A cell containing nucleic acid according to any of claims 1 to 9.

12. A cell according to claim 11 which is a muscle cell.

13. A cell according to claim 11 or claim 12 wherein said polypeptide is expressed.

14. A cell according to any of claims 11 to 13 which is in a mammal.

15. A mammal having a cell according to any of claims 11 to 13.

16. A mammal containing nucleic acid according to any of claims 1 to 9.

17. A method including introduction of nucleic acid according to any of claims 1 to 9 into a cell.

18. A method according to claim 17 wherein said cell is a muscle cell.

19. A method according to claim 17 or claim 18 wherein said introduction takes place *in vitro*.

20. A method which includes causing or allowing expression of the coding nucleotide sequence of nucleic acid according to any of claims 1 to 9 in a cell.

21. A method according to claim 20 wherein the cell is part of a mammal.

22. A method according to claim 20 wherein the expression product is purified and/or isolated following expression.

23. A method according to claim 22 wherein the expression product is formulated into a composition which includes at least one additional component, following purification and/or isolation of the expression product.

24. A polypeptide as encoded by nucleic acid according to any of claims 1 to 3.

25. A polypeptide as encoded by nucleic acid according to claim 4, excluding natural utrophin.

26. A composition including a polypeptide according to claim 24 or claim 25 and a pharmaceutically acceptable excipient.

27. A method for ameliorating one or more symptoms of a dystrophic phenotype in a mammal, the method including providing cells of the mammal with a polypeptide according to claim 24.

28. A method for ameliorating one or more symptoms of a dystrophic phenotype in a mammal, the method

including providing cells of the mammal with a polypeptide according to claim 25.

29. A method according to claim 27 or claim 28 wherein the polypeptide is provided to the cells by expression from encoding nucleic acid administered to the mammal.

30. Use of nucleic acid according to any of claims 1 to 9 in the manufacture of a medicament for treating a dystrophin phenotype in a mammal.

Figure 1a

Region within the actin binding which was unclonable in human.

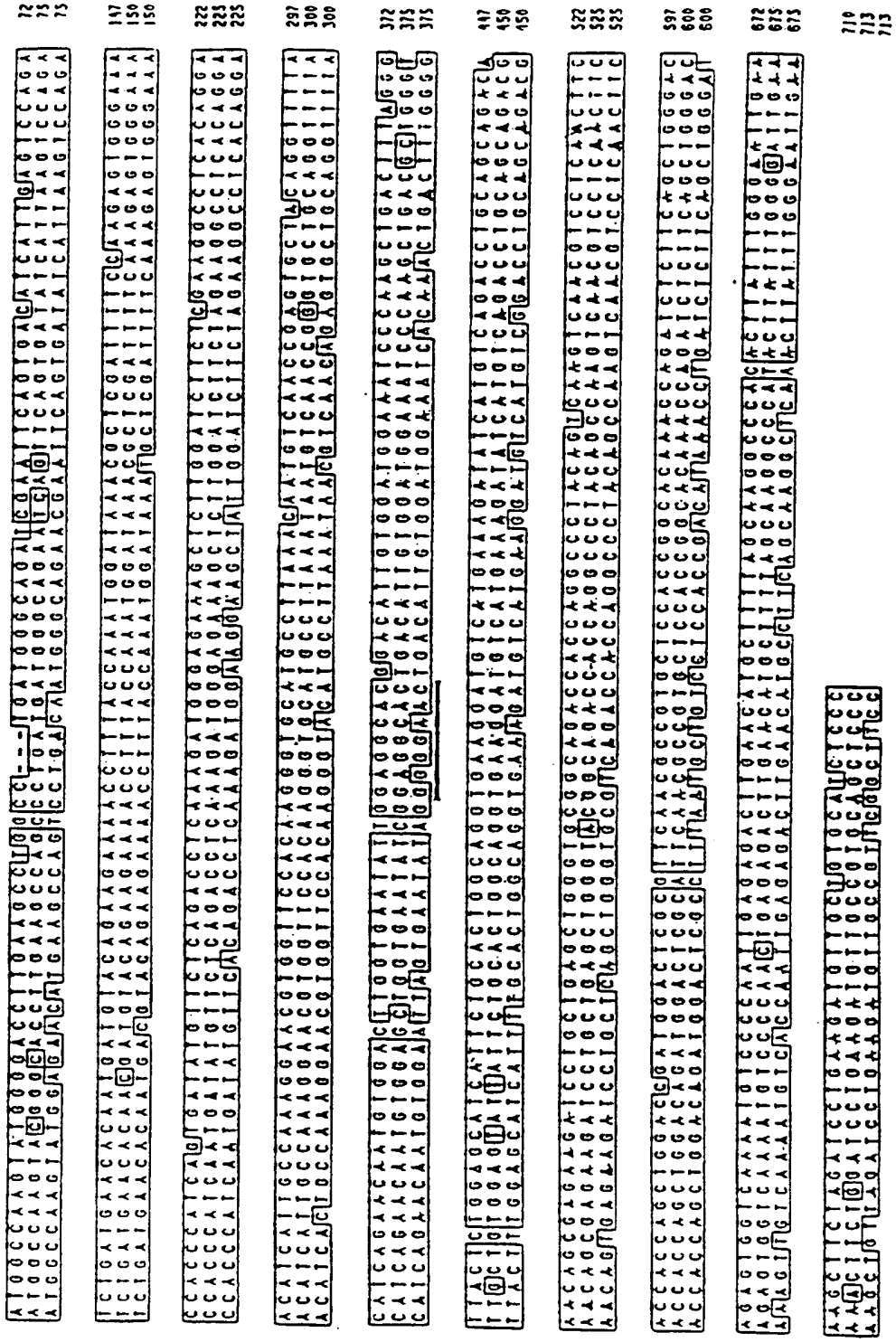


Figure 1b Sequence of 2nd unclonable region in human

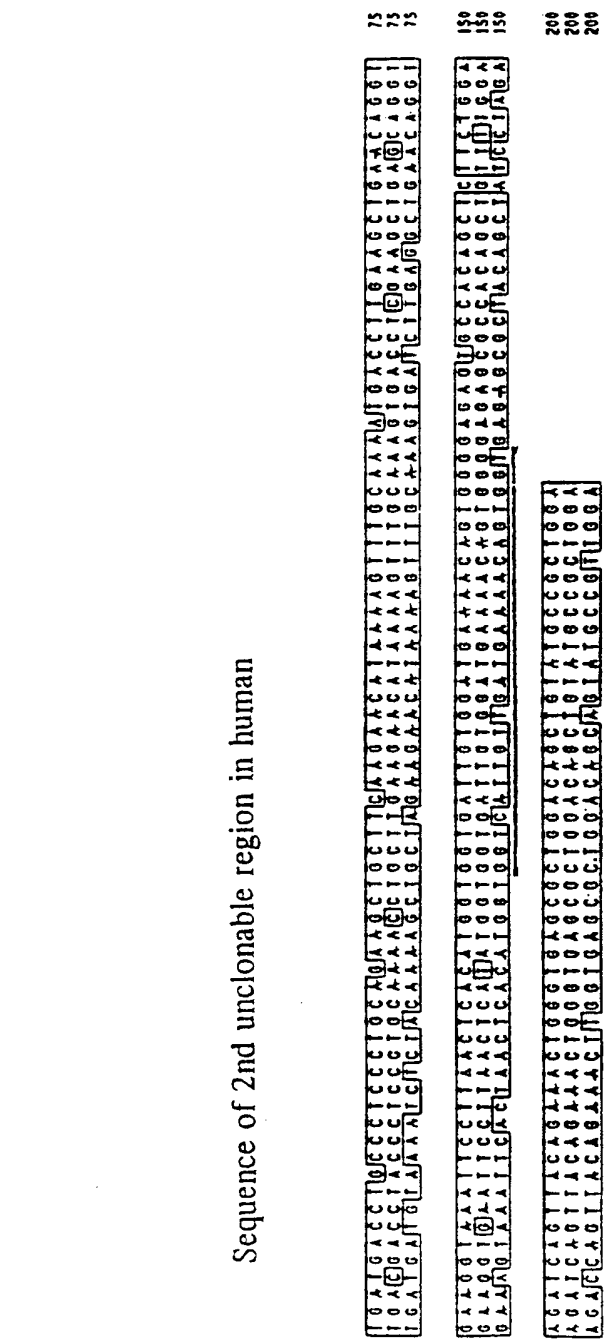


Figure 2

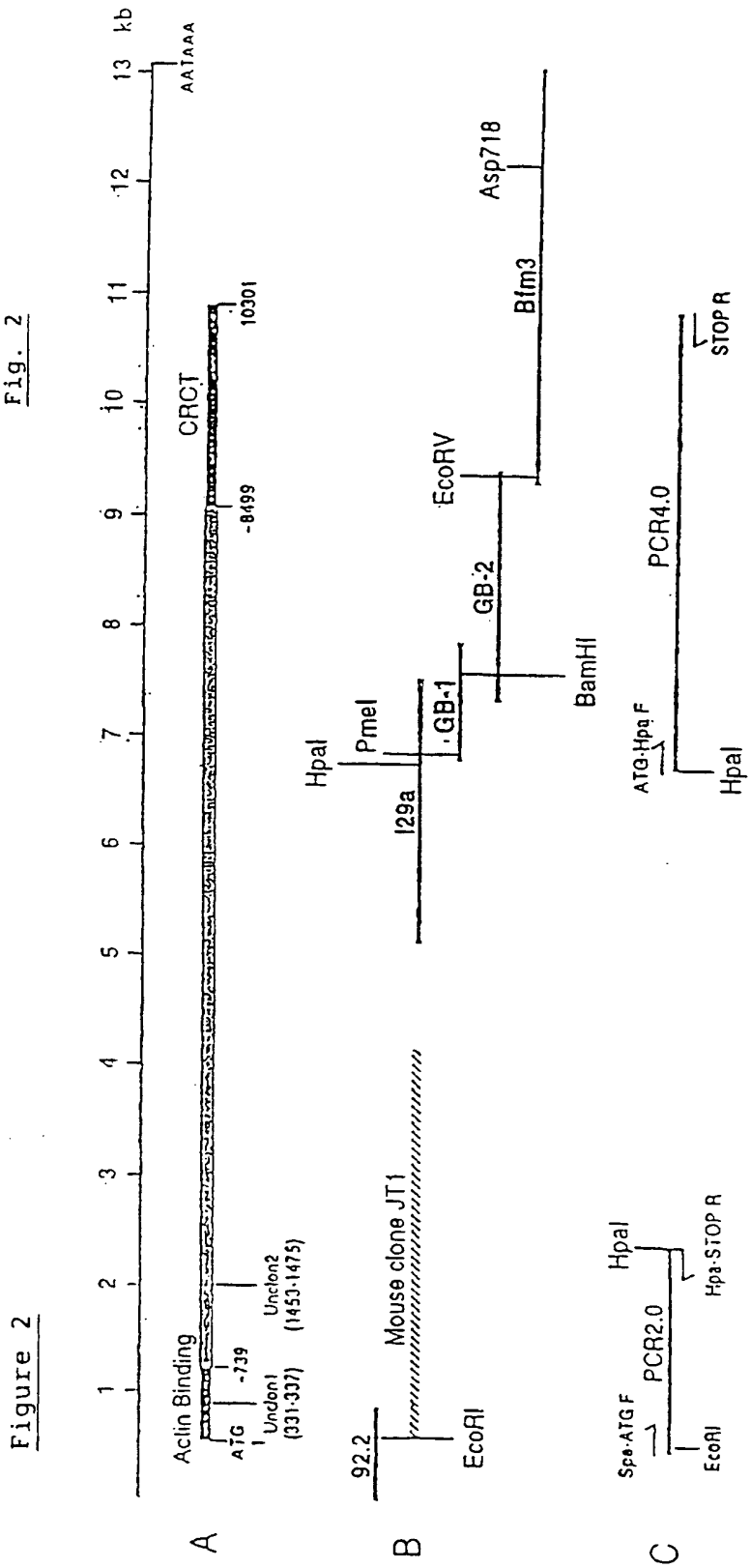


Fig. 2
1 2 3 4 5 6 7 8 9 10 11 12 13 kb
AATAAA
CRCT
10301
-8499
Actin Binding
ATG
-739
UndonI (331:337)
UndonI2 (1453:1475)
A
92.2
Mouse clone JT1
EcoRI
HpaI
PmeI
I29a
GB-1
GB-2
EcoRV
BamHI
Bfm3
Asp718
B
Spe-ATG F
PCR2.0
EcoRI
HpaI
Hpa-STOP R
ATG-Hpa F
PCR4.0
HpaI
STOP R
C

Figure 3

```
                M  A  K  Y  G  E  H  E  A  S  P  D  N  G  Q  N  E
   1 ACTAGTCAAGATGGCCAAGTATGGAGAACATGAAGCCAGTCCTGACAATGGGCAGAACGA   60

             F  S  D  I  I  E  S  R  S  D  E  H  N  D  V  Q  K  K  T  F
  61 ATTCAGTGACATCATTGAGTCCAGATCTGATGAACACAATGATGTACAGAAGAAAACCTT  120

             T  K  W  I  N  A  R  F  S  K  S  G  K  P  P  I  S  D  M  F
 121 TACCAAATGGATAAACGCTCGATTTTCCAAGAGTGGGAAACCACCCATCAGTGATATGTT  180

             S  D  L  K  D  G  R  K  L  L  D  L  L  E  G  L  T  G  T  S
 181 CTCAGACCTCAAAGATGGGAGAAAGCTCTTGGATCTTCTCGAAGGCCTCACAGGAACATC  240

             L  P  K  E  R  G  S  T  R  V  H  A  L  N  N  V  N  R  V  L
 241 ATTGCCAAAGGAACGTGGTTCCACAAGGGTGCATGCCTTAAACAATGTCAACCGAGTGCT  300

             Q  V  L  H  Q  N  N  V  D  L  V  N  I  G  G  T  D  I  V  D
 301 ACAGGTTTTACATCAGAACAATGTGGACTTGGTGAATATTGGAGGCACGGACATTGTGGA  360

             G  N  P  K  L  T  L  G  L  L  W  S  I  I  L  H  W  Q  V  K
 361 TGGAAATCCCAAGCTGACTTTAGGGTTACTCTGGAGCATCATTCTGCACTGGCAGGTGAA  420

             D  V  M  K  D  I  M  S  D  L  Q  Q  T  N  S  E  K  I  L  L
 421 GGATGTCATGAAAGATATCATGTCAGACCTGCAGCAGACAAACAGCGAGAAGATCCTGCT  480

             S  W  V  R  Q  T  T  R  P  Y  S  Q  V  N  V  L  N  F  T  T
 481 GAGCTGGGTGCGGCAGACCACCAGGCCCTACAGTCAAGTCAACGTCCTCAACTTCACCAC  540

             S  W  T  D  G  L  A  F  N  A  V  L  H  R  H  K  P  D  L  F
 541 CAGCTGGACCGATGGACTCGCGTTCAACGCCGTGCTCCACCGGCACAAACCAGATCTCTT  600

             S  W  D  R  V  V  K  M  S  P  I  E  R  L  E  H  A  F  S  K
 601 CAGCTGGGACAGAGTGGTCAAAATGTCCCCAATTGAGAGACTTGAACATGCTTTTAGCAA  660

             A  H  T  Y  L  G  I  E  K  L  L  D  P  E  D  V  A  V  H  L
 661 GGCCCACACTTATTTGGGAATTGAAAAGCTTCTAGATCCTGAAGATGTTGCTGTGCATCT  720

             P  X  X  X  X  X  X  X  X  X  X  X  X  V  E  V  L  P  Q  Q
 721 CCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCGTTGAGGTGCTTCCTCAGCA  780

             V  T  I  D  A  I  R  E  V  E  T  L  P  R  K  Y  K  K  E  C
 781 AGTCACGATAGATGCCATCCGAGAGGTGGAGACTCTCCCAAGGAAGTATAAGAAAGAATG  840

             E  E  E  E  I  H  I  Q  S  A  V  L  A  E  E  G  Q  S  P  R
 841 TGAAGAGGAAGAAATTCATATCCAGAGTGCAGTGCTGGCAGAGGAAGGCCAGAGTCCCCG  900

             A  E  T  P  S  T  V  T  E  V  D  M  D  L  D  S  Y  Q  I  A
 901 AGCTGAGACCCCTAGCACCGTCACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGC  960

             L  E  E  V  L  T  W  L  L  S  A  E  D  T  F  Q  E  Q  D  D
 961 GCTAGAGGAAGTGCTGACGTGGCTGCTGTCCGCGGAGGACACGTTCCAGGAGCAAGATGA 1020

             I  S  D  D  V  E  E  V  K  E  Q  F  A  T  H  E  T  F  M  M
1021 CATTTCTGATGATGTCGAAGAAGTCAAAGAGCAGTTTGCTACCCATGAAACTTTTATGAT 1080

             E  L  T  A  H  Q  S  S  V  G  S  V  L  Q  A  G  N  Q  L  M
1081 GGAGCTGACAGCACACCAGAGCAGCGTGGGGAGCGTCCTGCAGGCTGGCAACCAGCTGAT 1140

             T  Q  G  T  L  S  E  E  E  E  F  E  I  Q  E  Q  M  T  L  L
1141 GACACAAGGGACTCTGTCAGAGGAGGAGGAGTTTGAGATCCAGGAACAGATGACCTTGCT 1200
```

Figure 3 Continued


```
          N  A  R  W  E  A  L  R  V  E  S  M  E  R  Q  S  R  L  H  D
1201  GAATGCAAGGTGGGAGGCGCTCCGGGTGGAGAGCATGGAGAGGCAGTCCCGGCTGCACGA  1260

          A  L  M  E  L  Q  K  K  Q  L  Q  Q  L  S  S  W  L  A  L  T
1261  CGCTCTGATGGAGCTGCAGAAGAAACAGCTGCAGCAGCTCTCAAGCTGGCTGGCCCTCAC  1320

          E  E  R  I  Q  K  M  E  S  P  P  L  G  D  D  L  P  S  L  Q
1321  AGAAGAGCGCATTCAGAAGATGGAGAGCCCTCCGCTGGGTGATGACCTGCCCTCCCTGCA  1380

          K  L  L  Q  E  H  K  S  L  Q  N  D  L  E  A  E  Q  V  K  V
1381  GAAGCTGCTTCAAGAACATAAAAGTTTGCAAAATGACCTTGAAGCTGAACAGGTGAAGGT  1440

          N  S  L  T  H  M  V  V  I  V  D  E  N  S  G  E  S  A  T  A
1441  AAATTCCTTAACTCACATGGTGGTGATTGTGGATGAAAACAGTGGGGAGAGTGCCACAGC  1500

          L  L  E  D  Q  L  Q  K  L  G  E  R  W  T  A  V  C  R  W  T
1501  TCTTCTGGAAGATCAGTTACAGAAACTGGGTGAGCGCTGGACAGCTGTATGCCGCTGGAC  1560

          E  E  R  W  N  R  L  Q  E  I  S  I  L  W  Q  E  L  L  E  E
1561  TGAAGAACGTTGGAACAGGTTGCAAGAAATCAGTATTCTGTGGCAGGAATTATTGGAAGA  1620

          Q  C  L  L  E  A  W  L  T  E  K  E  E  A  L  N  K  V  Q  T
1621  GCAGTGTCTGTTGGAGGCTTGGCTCACCGAAAAGGAAGAGGCTTTGAATAAAGTTCAAAC  1680

          S  N  F  K  D  Q  K  E  L  S  V  S  V  R  R  L  A  I  L  K
1681  CAGCAACTTTAAAGACCAGAAGGAACTAAGTGTCAGTGTCCGGCGTCTGGCTATATTGAA  1740

          E  D  M  E  M  K  R  Q  T  L  D  Q  L  S  E  I  G  Q  D  V
1741  GGAAGACATGGAAATGAAGAGGCAGACTCTGGATCAACTGAGTGAGATTGGCCAGGATGT  1800

          G  Q  L  L  S  N  P  K  A  S  K  K  M  N  S  D  S  E  E  L
1801  GGGCCAATTACTCAGTAATCCCAAGGCATCTAAGAAGATGAACAGTGACTCTGAGGAGCT  1860

          T  Q  R  W  D  S  L  V  Q  R  L  E  D  S  S  N  Q  V  T  Q
1861  AACACAGAGATGGGATTCTCTGGTTCAGAGACTCGAAGACTCTTCTAACCAGGTGACTCA  1920

          A  V  A  K  L  G  M  S  Q  I  P  Q  K  D  L  L  E  T  V  H
1921  GGCGGTAGCGAAGCTCGGCATGTCCCAGATTCCACAGAAGGACCTATTGGAGACCGTTCA  1980

          V  R  E  K  G  M  V  K  K  P  K  Q  E  L  P  P  P  L  T  K
1981  TGTGAGAGAAAAAGGGATGGTGAAGAAGCCCAAGCAGGAACTGCCTCCTCCGTTAACAAA  2040

          A  E  H  A  M  Q  K  R  S  T  T  E  L  G  E  N  L  Q  E  L
2041  GGCTGAGCATGCTATGCAAAAGAGATCAACCACCGAATTGGGAGAAAACCTGCAAGAATT  2100

          R  D  L  T  Q  E  M  E  V  H  A  E  K  L  K  W  L  N  R  T
2101  AAGAGACTTAACTCAAGAAATGGAAGTACATGCTGAAAAACTCAAATGGCTGAATAGAAC  2160

          E  L  E  M  L  S  D  K  S  L  S  L  P  E  R  D  K  I  S  E
2161  TGAATTGGAGATGCTTTCAGATAAAAGTCTGAGTTTACCTGAAAGGGATAAAATTTCAGA  2220

          S  L  R  T  V  N  M  T  W  N  K  I  C  R  E  V  P  T  T  L
2221  AAGCTTAAGGACTGTAAATATGACATGGAATAAGATTTGCAGAGAGGTGCCTACCACCCT  2280

          K  E  C  I  Q  E  P  S  S  V  S  Q  T  R  I  A  A  H  P  N
2281  GAAGGAATGCATCCAGGAGCCCAGTTCTGTTTCACAGACAAGGATTGCTGCTCATCCTAA  2340

          V  Q  K  V  V  L  V  S  S  A  S  D  I  P  V  Q  S  H  R  T
2341  TGTCCAAAAGGTGGTGCTAGTATCATCTGCGTCAGATATTCCTGTTCAGTCTCATCGTAC  2400

          S  E  I  S  I  P  A  D  L  D  K  T  I  T  E  L  A  D  W  L
2401  TTCGGAAATTTCAATTCCTGCTGATCTTGATAAAACTATAACAGAACTAGCCGACTGGCT  2460
```

Figure 3 Continued

```
      V  L  I  D  Q  M  L  K  S  N  I  V  T  V  G  D  V  E  E  I
2461  GGTATTAATCGACCAGATGCTGAAGTCCAACATTGTCACTGTTGGGGATGTAGAAGAGAT  2520

      N  K  T  V  S  R  M  K  I  T  K  A  D  L  E  Q  R  H  P  Q
2521  CAATAAGACCGTTTCCCGAATGAAAATTACAAAGGCTGACTTAGAACAGCGCCATCCTCA  2580

      L  D  Y  V  F  T  L  A  Q  N  L  K  N  K  A  S  S  S  D  M
2581  GCTGGATTATGTTTTTACATTGGCACAGAATTTGAAAAATAAAGCTTCCAGTTCAGATAT  2640

      R  T  A  I  T  E  K  L  E  R  V  K  N  Q  W  D  G  T  Q  H
2641  GAGAACAGCAATTACAGAAAAATTGGAAAGGGTCAAGAACCAGTGGGATGGCACCCAGCA  2700

      G  V  E  L  R  Q  Q  Q  L  E  D  M  I  I  D  S  L  Q  W  D
2701  TGGCGTTGAGCTAAGACAGCAGCAGCTTGAGGACATGATTATTGACAGTCTTCAGTGGGA  2760

      D  H  R  E  E  T  E  E  L  M  R  K  Y  E  A  R  L  Y  I  L
2761  TGACCATAGGGAGGAGACTGAAGAACTGATGAGAAAATATGAGGCTCGACTCTATATTCT  2820

      Q  Q  A  R  R  D  P  L  T  K  Q  I  S  D  N  Q  I  L  L  Q
2821  TCAGCAAGCCCGACGGGATCCACTCACCAAACAAATTTCTGATAACCAAATACTGCTTCA  2880

      E  L  G  P  G  D  G  I  V  M  A  F  D  N  V  L  Q  K  L  L
2881  AGAACTGGGTCCTGGAGATGGTATCGTCATGGCGTTCGATAACGTCCTGCAGAAACTCCT  2940

      E  E  Y  G  S  D  D  T  R  N  V  K  E  T  T  E  Y  L  K  T
2941  GGAGGAATATGGGAGTGATGACACAAGGAATGTGAAAGAAACCACAGAGTACTTAAAAAC  3000

      S  W  I  N  L  K  Q  S  I  A  D  R  Q  N  A  L  E  A  E  W
3001  ATCATGGATCAATCTCAAACAAAGTATTGCTGACAGACAGAACGCCTTGGAGGCTGAGTG  3060

      R  T  V  Q  A  S  R  R  D  L  E  N  F  L  K  W  I  Q  E  A
3061  GAGGACGGTGCAGGCCTCTCGCAGAGATCTGGAAAACTTCCTGAAGTGGATCCAAGAAGC  3120

      E  T  T  V  N  V  L  V  D  A  S  H  R  E  N  A  L  Q  D  S
3121  AGAGACCACAGTGAATGTGCTTGTGGATGCCTCTCATCGGGAGAATGCTCTTCAGGATAG  3180

      I  L  A  R  E  L  K  Q  Q  M  Q  D  I  Q  A  E  I  D  A  H
3181  TATCTTGGCCAGGGAACTCAAACAGCAGATGCAGGACATCCAGGCAGAAATTGATGCCCA  3240

      N  D  I  F  K  S  I  D  G  N  R  Q  K  M  V  K  A  L  G  N
3241  CAATGACATATTTAAAAGCATTGACGGAAACAGGCAGAAGATGGTAAAAGCTTTGGGAAA  3300

      S  E  E  A  T  M  L  Q  H  R  L  D  D  M  N  Q  R  W  N  D
3301  TTCTGAAGAGGCTACTATGCTTCAACATCGACTGGATGATATGAACCAAAGATGGAATGA  3360

      L  K  A  K  S  A  S  I  R  A  H  L  E  A  S  A  E  K  W  N
3361  CTTAAAAGCAAAATCTGCTAGCATCAGGGCCCATTTGGAGGCCAGCGCTGAGAAGTGGAA  3420

      R  L  L  M  S  L  E  E  L  I  K  W  L  N  M  K  D  E  E  L
3421  CAGGTTGCTGATGTCCTTAGAAGAACTGATCAAATGGCTGAATATGAAAGATGAAGAGCT  3480

      K  K  Q  M  P  I  G  G  D  V  P  A  L  Q  L  Q  Y  D  H  C
3481  TAAGAAACAAATGCCTATTGGAGGAGATGTTCCAGCCTTACAGCTCCAGTATGACCATTG  3540

      K  A  L  R  R  E  L  K  E  K  E  Y  S  V  L  N  A  V  D  Q
3541  TAAGGCCCTGAGACGGGAGTTAAAGGAGAAAGAATATTCTGTCCTGAATGCTGTCGACCA  3600

      A  R  V  F  L  A  D  Q  P  I  E  A  P  E  E  P  R  R  N  L
3601  GGCCCGAGTTTTCTTGGCTGATCAGCCAATTGAGGCCCCTGAAGAGCCAAGAAGAAACCT  3660

      Q  S  K  T  E  L  T  P  E  E  R  A  Q  K  I  A  K  A  M  R
3661  ACAATCAAAAACAGAATTAACTCCTGAGGAGAGAGCCCAAAAGATTGCCAAAGCCATGCG  3720
```

Figure 3 Continued

```
          K  Q  S  S  E  V  K  E  K  W  E  S  L  N  A  V  T  S  N  W
3721  CAAACAGTCTTCTGAAGTCAAAGAAAAATGGGAAAGTCTAAATGCTGTAACTAGCAATTG  3780

          Q  K  Q  V  D  K  A  L  E  K  L  R  D  L  Q  G  A  M  D  D
3781  GCAAAAGCAAGTGGACAAGGCATTGGAGAAACTCAGAGACCTGCAGGGAGCTATGGATGA  3840

          L  D  A  D  M  K  E  A  E  S  V  R  N  G  W  K  P  V  G  D
3841  CCTGGACGCTGACATGAAGGAGGCAGAGTCCGTGCGGAATGGCTGGAAGCCCGTGGGAGA  3900

          L  L  I  D  S  L  Q  D  H  I  E  K  I  M  A  F  R  E  E  I
3901  CTTACTCATTGACTCGCTGCAGGATCACATTGAAAAAATCATGGCATTTAGAGAAGAAAT  3960

          A  P  I  N  F  K  V  K  T  V  N  D  L  S  S  Q  L  S  P  L
3961  TGCACCAATCAACTTTAAAGTTAAAACGGTGAATGATTTATCCAGTCAGCTGTCTCCACT  4020

          D  L  H  P  S  L  K  M  S  R  Q  L  D  D  L  N  M  R  W  K
4021  TGACCTGCATCCCTCTCTAAAGATGTCTCGCCAGCTAGATGACCTTAATATGCGATGGAA  4080

          L  L  Q  V  S  V  D  D  R  L  K  Q  L  Q  E  A  H  R  D  F
4081  ACTTTTACAGGTTTCTGTGGATGATCGCCTTAAACAGCTTCAGGAAGCCCACAGAGATTT  4140

          G  P  S  S  Q  H  F  L  S  T  S  V  Q  L  P  W  Q  R  S  I
4141  TGGACCATCCTCTCAGCATTTTCTCTCTACGTCAGTCCAGCTGCCGTGGCAAAGATCCAT  4200

          S  H  N  K  V  P  Y  Y  I  N  H  Q  T  Q  T  T  C  W  D  H
4201  TTCACATAATAAAGTGCCCTATTACATCAACCATCAAACACAGACCACCTGTTGGGACCA  4260

          P  K  M  T  E  L  F  Q  S  L  A  D  L  N  N  V  R  F  S  A
4261  TCCTAAAATGACCGAACTCTTTCAATCCCTTGCTGACCTGAATAATGTACGTTTTTCTGC  4320

          Y  R  T  A  I  K  I  R  R  L  Q  K  A  L  C  L  D  L  L  E
4321  CTACCGTACAGCAATCAAAATCCGAAGACTACAAAAAGCACTATGTTTGGATCTCTTAGA  4380

          L  S  T  T  N  E  I  F  K  Q  H  K  L  N  Q  N  D  Q  L  L
4381  GTTGAGTACAACAAATGAAATTTTCAAACAGCACAAGTTGAACCAAAATGACCAGCTCCT  4440

          S  V  P  D  V  I  N  C  L  T  T  T  Y  D  G  L  E  Q  M  H
4441  CAGTGTTCCAGATGTCATCAACTGTCTGACAACAACTTATGATGGACTTGAGCAAATGCA  4500

          K  D  L  V  N  V  P  L  C  V  D  M  C  L  N  W  L  L  N  V
4501  TAAGGACCTGGTCAACGTTCCACTCTGTGTTGATATGTGTCTCAATTGGTTGCTCAATGT  4560

          Y  D  T  G  R  T  G  K  I  R  V  Q  S  L  K  I  G  L  M  S
4561  CTATGACACGGGTCGAACTGGAAAAATTAGAGTGCAGAGTCTGAAGATTGGATTAATGTC  4620

          L  S  K  G  L  L  E  E  K  Y  R  Y  L  F  K  E  V  A  G  P
4621  TCTCTCCAAAGGTCTCTTGGAAGAAAAATACAGATATCTCTTTAAGGAAGTTGCGGGGCC  4680

          T  E  M  C  D  Q  R  Q  L  G  L  L  L  H  D  A  I  Q  I  P
4681  GACAGAAATGTGTGACCAGAGGCAGCTGGGCCTGTTACTTCATGATGCCATCCAGATCCC  4740

          R  Q  L  G  E  V  A  A  F  G  G  S  N  I  E  P  S  V  R  S
4741  CCGGCAGCTAGGTGAAGTAGCAGCTTTTGGAGGCAGTAATATTGAGCCTAGTGTTCGCAG  4800

          C  F  Q  Q  N  N  N  K  P  E  I  S  V  K  E  F  I  D  W  M
4801  CTGCTTCCAACAGAATAACAATAAACCAGAAATAAGTGTGAAAGAGTTTATAGATTGGAT  4860

          H  L  E  P  Q  S  M  V  W  L  P  V  L  H  R  V  A  A  A  E
4861  GCATTTGGAACCACAGTCCATGGTTTGGCTCCCAGTTTTACATCGAGTGGCAGCAGCGGA  4920

          T  A  K  H  Q  A  K  C  N  I  C  K  E  C  P  I  V  G  F  R
4921  GACTGCAAAACATCAGGCCAAATGCAACATCTGTAAAGAATGTCCAATTGTCGGGTTCAG  4980
```

Figure 3 Continued

```
          Y  R  S  L  K  H  F  N  Y  D  V  C  Q  S  C  F  F  S  G  R
4981  GTATAGAAGCCTTAAGCATTTTAACTATGATGTCTGCCAGAGTTGTTTCTTTTCGGGTCG  5040

          T  A  K  G  H  K  L  H  Y  P  M  V  E  Y  C  I  P  T  T  S
5041  AACAGCAAAAGGTCACAAATTACATTACCCAATGGTGGAATATTGTATACCTACAACATC  5100

          G  E  D  V  R  D  F  T  K  V  L  K  N  K  F  R  S  K  K  Y
5101  TGGGGAAGATGTACGAGACTTCACAAAGGTACTTAAGAACAAGTTCAGGTCGAAGAAGTA  5160

          F  A  K  H  P  R  L  G  Y  L  P  V  Q  T  V  L  E  G  D  N
5161  CTTTGCCAAACACCCTCGACTTGGTTACCTGCCTGTCCAGACAGTTCTTGAAGGTGACAA  5220

          L  E  T  P  I  T  L  I  S  M  W  P  E  H  Y  D  P  S  Q  S
5221  CTTAGAGACTCCTATCACACTCATCAGTATGTGGCCAGAGCACTATGACCCCTCACAATC  5280

          P  Q  L  F  H  D  D  T  H  S  R  I  E  Q  Y  A  T  R  L  A
5281  TCCTCAACTGTTTCATGATGACACCCATTCAAGAATAGAACAATATGCCACACGACTGGC  5340

          Q  M  E  R  T  N  G  S  F  L  T  D  S  S  S  T  T  G  S  V
5341  CCAGATGGAAAGGACTAATGGGTCTTTTCTCACTGATAGCAGCTCCACCACAGGAAGTGT  5400

          E  D  E  H  A  L  I  Q  Q  Y  C  Q  T  L  G  G  E  S  P  V
5401  GGAAGACGAGCACGCCCTCATCCAGCAGTATTGCCAAACACTCGGAGGAGAGTCCCCAGT  5460

          S  Q  P  Q  S  P  A  Q  I  L  K  S  V  E  R  E  E  R  G  E
5461  GAGCCAGCCGCAGAGCCCAGCTCAGATCCTGAAGTCAGTAGAGAGGGAAGAACGTGGAGA  5520

          L  E  R  I  I  A  D  L  E  E  E  Q  R  N  L  Q  V  E  Y  E
5521  ACTGGAGAGGATCATTGCTGACCTGGAGGAAGAACAAAGAAATCTACAGGTGGAGTATGA  5580

          Q  L  K  D  Q  H  L  R  R  G  L  P  V  G  S  P  P  E  S  I
5581  GCAGCTGAAGGACCAGCACCTCCGAAGGGGGCTCCCTGTCGGTTCACCGCCAGAGTCGAT  5640

          I  S  P  H  H  T  S  E  D  S  E  L  I  A  E  A  K  L  L  R
5641  TATATCTCCCCATCACACGTCTGAGGATTCAGAACTTATAGCAGAAGCAAAACTCCTCAG  5700

          Q  H  K  G  R  L  E  A  R  M  Q  I  L  E  D  H  N  K  Q  L
5701  GCAGCACAAAGGTCGGCTGGAGGCTAGGATGCAGATTTTAGAAGATCACAATAAACAGCT  5760

          E  S  Q  L  H  R  L  R  Q  L  L  E  Q  P  E  S  D  S  R  I
5761  GGAGTCTCAGCTCCACCGCCTCCGACAGCTGCTGGAGCAGCCTGAATCTGATTCCCGAAT  5820

          N  G  V  S  P  W  A  S  P  Q  H  S  A  L  S  Y  S  L  D  P
5821  CAATGGTGTTTCCCCATGGGCTTCTCCTCAGCATTCTGCACTGAGCTACTCGCTTGATCC  5880

          D  A  S  G  P  Q  F  H  Q  A  A  G  E  D  L  L  A  P  P  H
5881  AGATGCCTCCGGCCCACAGTTCCACCAGGCAGCGGGAGAGGACCTGCTGGCCCCACCGCA  5940

          D  T  S  T  D  L  T  E  V  M  E  Q  I  H  S  T  F  P  S  C
5941  CGACACCAGCACGGATCTCACGGAGGTCATGGAGCAGATTCACAGCACGTTTCCATCTTG  6000

          C  P  N  V  P  S  R  P  Q  A  M  *
6001  CTGCCCAAATGTTCCCAGCAGGCCACAGGCAATGTAATCACTAGT  6045
```

Figure 4

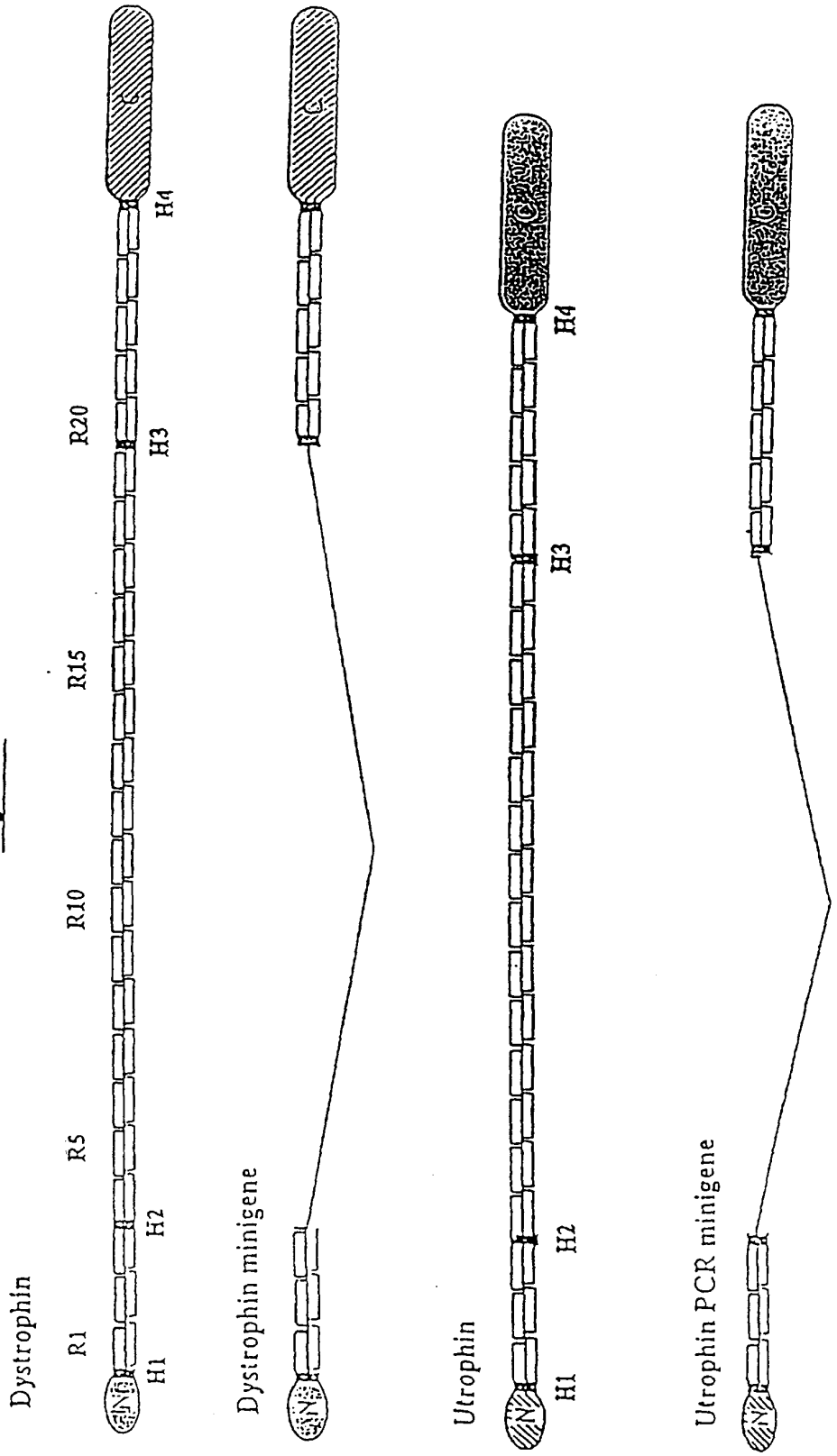
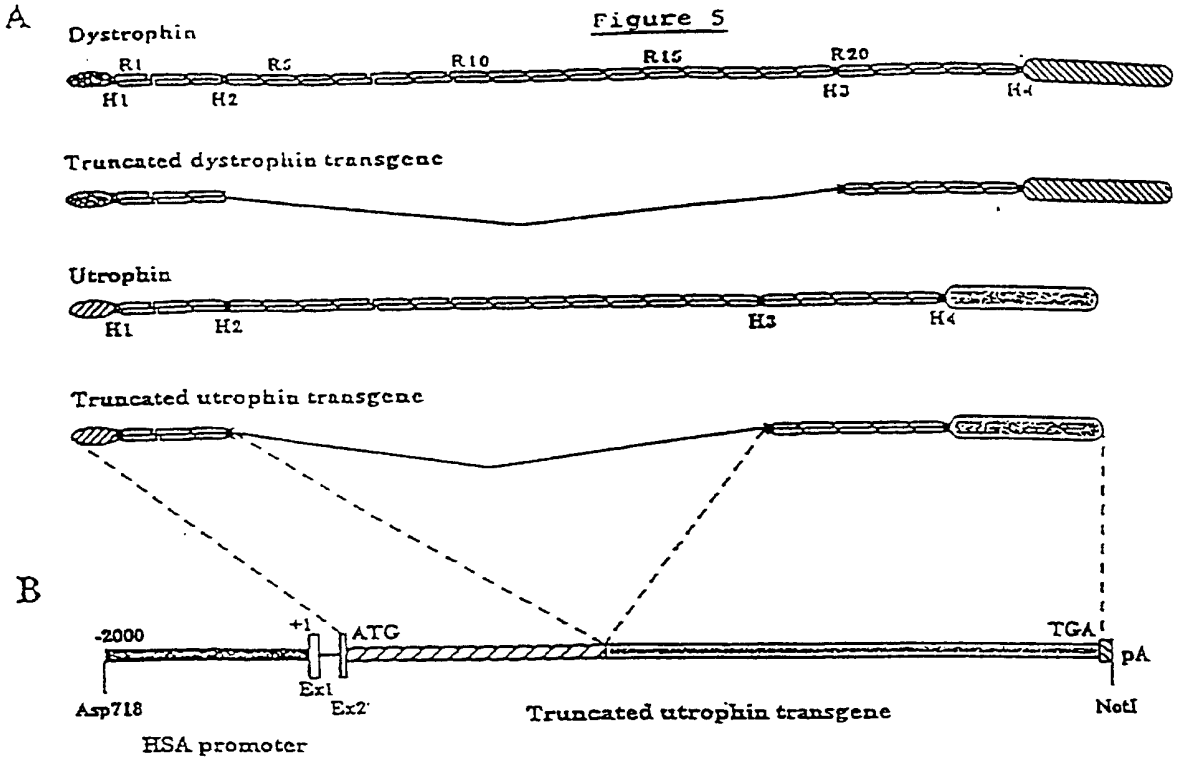


Figure 5



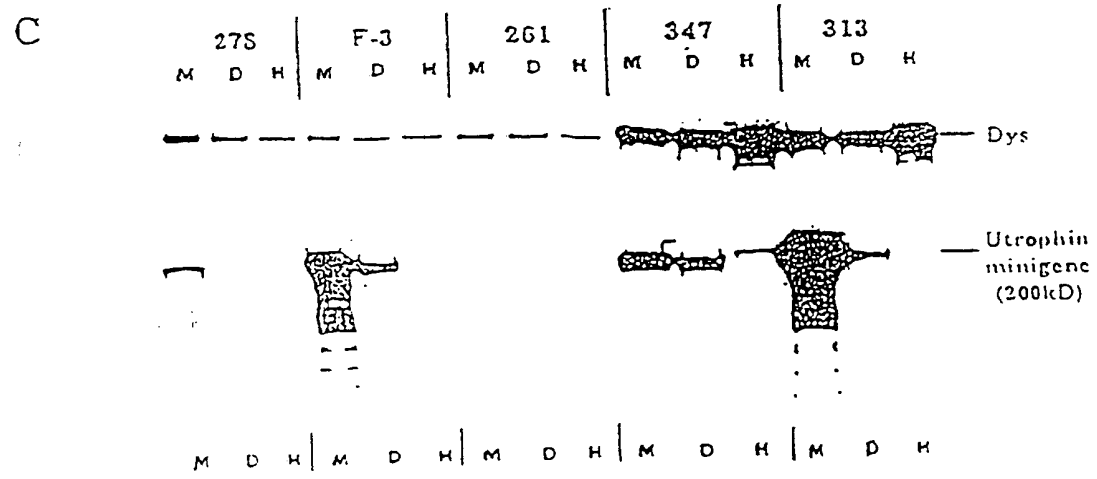


Figure 6

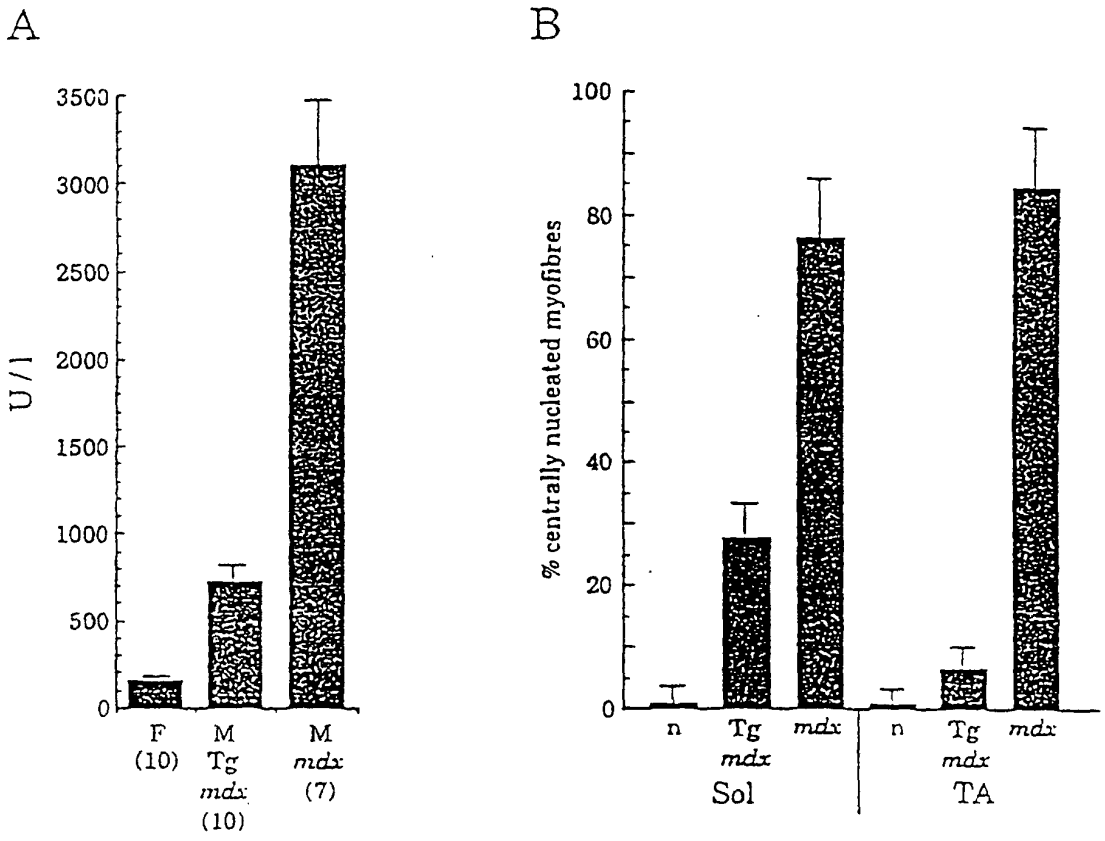
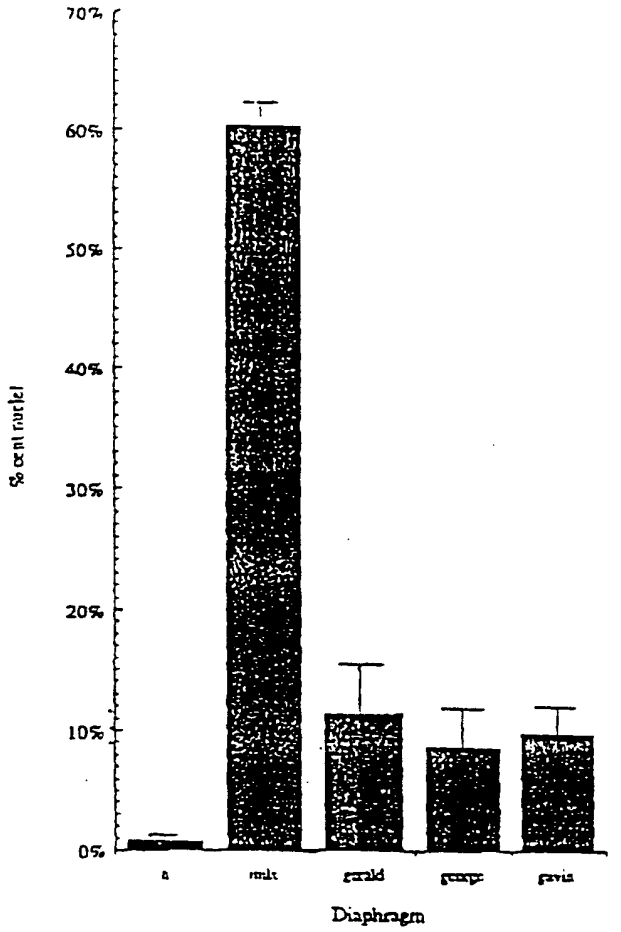

A
U/l
3500
3000
2500
2000
1500
1000
500
0
F (10)
M Tg mdx (10)
M mdx (7)
B
% centrally nucleated myofibres
100
80
60
40
20
0
n
Tg mdx
mdx
Sol
n
Tg mdx
mdx
TA

Figure 7


70%
60%
50%
40%
30%
20%
10%
0%
% central nuclei
Diaphragm

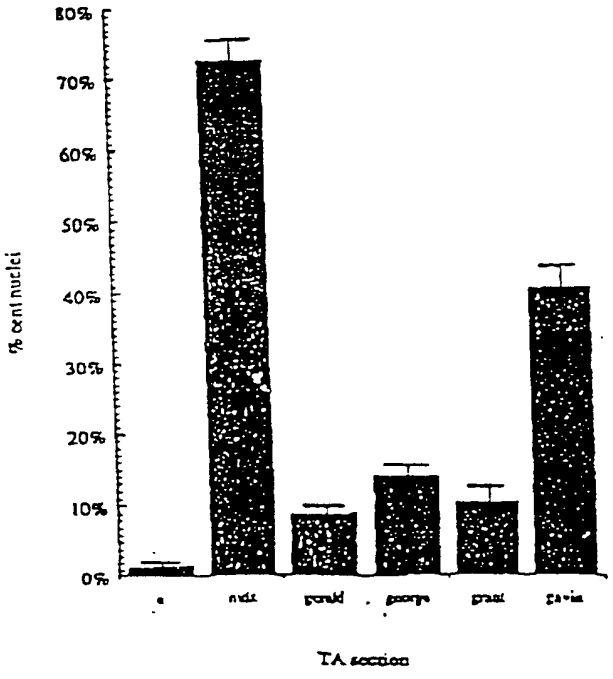

80%
70%
60%
50%
40%
30%
20%
10%
0%
% central nuclei
TA section

Figure 8

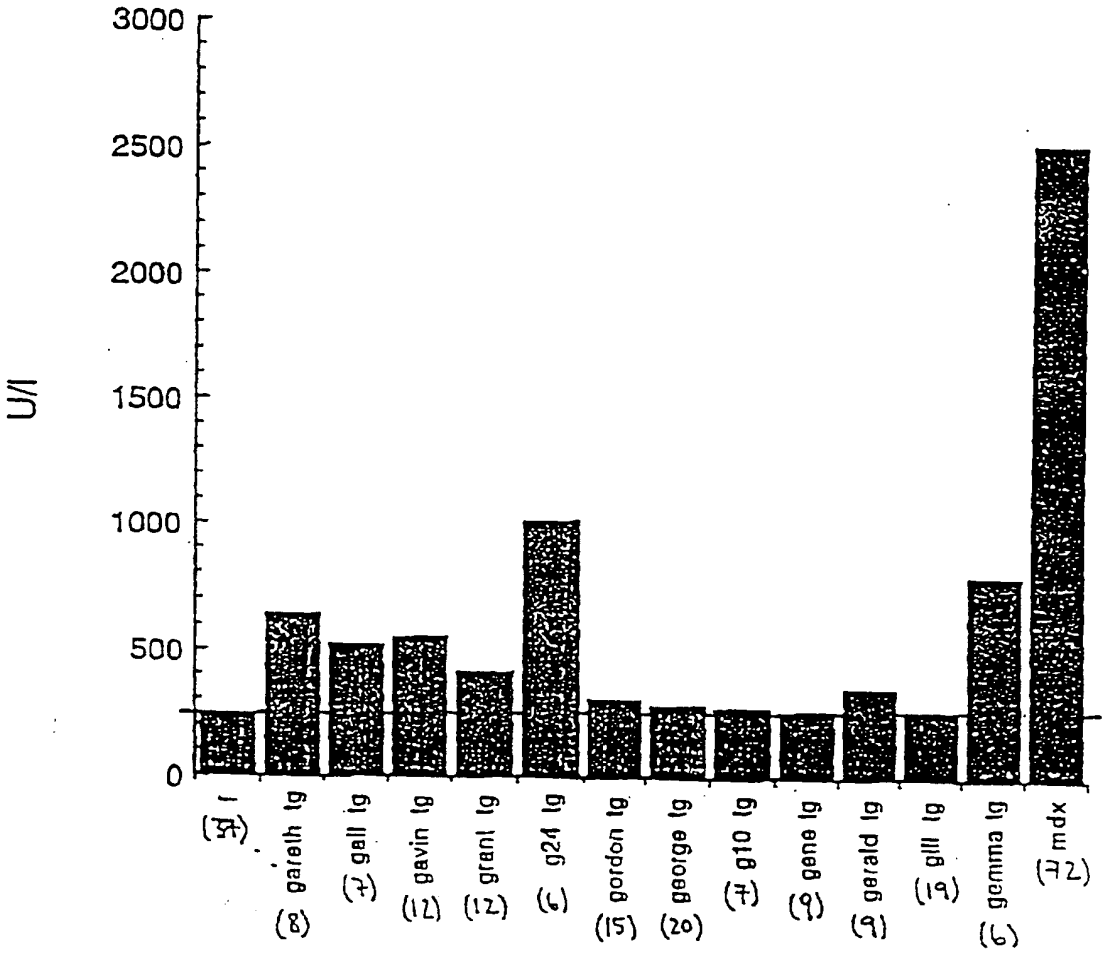

3000
2500
2000
1500
1000
500
0
U/l
r (5)
gareth Ig (8)
gall Ig (7)
gavin Ig (12)
granl Ig (12)
g24 Ig (6)
gordon Ig (15)
george Ig (20)
g10 Ig (7)
gene Ig (9)
gerald Ig (9)
gill Ig (19)
gemma Ig (6)
mdx (72)

Figure 9

```
                M  A  K  Y  G  E  H  E  A  S  P  D  N  G  Q  N  E
  1 ACTAGTCAAGATGGCCAAGTATGGAGAACATGAAGCCAGTCCTGACAATGGGCAGAACGA  60

             F  S  D  I  I  E  S  R  S  D  E  H  N  D  V  Q  K  K  T  F
 61 ATTCAGTGACATCATTGAGTCCAGATCTGATGAACACAATGATGTACAGAAGAAAACCTT 120

          T  K  W  I  N  A  R  F  S  K  S  G  K  P  P  I  S  D  M  F
121 TACCAAATGGATAAACGCTCGATTTTCCAAGAGTGGGAAACCACCCATCAGTGATATGTT 180

       S  D  L  K  D  G  R  K  L  L  D  L  L  E  G  L  T  G  T  S
181 CTCAGACCTCAAAGATGGGAGAAAGCTCTTGGATCTTCTCGAAGGCCTCACAGGAACATC 240

    L  P  K  E  R  G  S  T  R  V  H  A  L  N  N  V  N  R  V  L
241 ATTGCCAAAGGAACGTGGTTCCACAAGGGTGCATGCCTTAAACAATGTCAACCGAGTGCT 300

 Q  V  L  H  Q  N  N  V  D  L  V  N  I  G  G  T  D  I  V  D
301 ACAGGTTTTACATCAGAACAATGTGGACTTGGTGAATATTGGAGGCACGGACATTGTGGA 360

    G  N  P  K  L  T  L  G  L  L  W  S  I  I  L  H  W  Q  V  K
361 TGGAAATCCCAAGCTGACTTTAGGGTTACTCTGGAGCATCATTCTGCACTGGCAGGTGAA 420

    D  V  M  K  D  I  M  S  D  L  Q  Q  T  N  S  E  K  I  L  L
421 GGATGTCATGAAAGATATCATGTCAGACCTGCAGCAGACAAACAGCGAGAAGATCCTGCT 480

    S  W  V  R  Q  T  T  R  P  Y  S  Q  V  N  V  L  N  F  T  T
481 GAGCTGGGTGCGGCAGACCACCAGGCCCTACAGTCAAGTCAACGTCCTCAACTTCACCAC 540

    S  W  T  D  G  L  A  F  N  A  V  L  H  R  H  K  P  D  L  F
541 CAGCTGGACCGATGGACTCGCGTTCAACGCCGTGCTCCACCGGCACAAACCAGATCTCTT 600

    S  W  D  R  V  V  K  M  S  P  I  E  R  L  E  H  A  F  S  K
601 CAGCTGGGACAGAGTGGTCAAAATGTCCCCAATTGAGAGACTTGAACATGCTTTTAGCAA 660

    A  H  T  Y  L  G  I  E  K  L  L  D  P  E  D  V  A  V  H  L
661 GGCCCACACTTATTTGGGAATTGAAAAGCTTCTAGATCCTGAAGATGTTGCTGTGCATCT 720

    P  X  X  X  X  X  X  X  X  X  X  X  X  V  E  V  L  P  Q  Q
721 CCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCGTTGAGGTGCTTCCTCAGCA 780

    V  T  I  D  A  I  R  E  V  E  T  L  P  R  K  Y  K  K  E  C
781 AGTCACGATAGATGCCATCCGAGAGGTGGAGACTCTCCCAAGGAAGTATAAGAAAGAATG 840

    E  E  E  E  I  H  I  Q  S  A  V  L  A  E  E  G  Q  S  P  R
841 TGAAGAGGAAGAAATTCATATCCAGAGTGCAGTGCTGGCAGAGGAAGGCCAGAGTCCCCG 900

    A  E  T  P  S  T  V  T  E  V  D  M  D  L  D  S  Y  Q  I  A
901 AGCTGAGACCCCTAGCACCGTCACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGC 960

    L  E  E  V  L  T  W  L  L  S  A  E  D  T  F  Q  E  Q  D  D
961 GCTAGAGGAAGTGCTGACGTGGCTGCTGTCCGCGGAGGACACGTTCCAGGAGCAAGATGA 1020

    I  S  D  D  V  E  E  V  K  E  Q  F  A  T  H  E  T  F  M  M
1021 CATTTCTGATGATGTCGAAGAAGTCAAAGAGCAGTTTGCTACCCATGAAACTTTTATGAT 1080

    E  L  T  A  H  Q  S  S  V  G  S  V  L  Q  A  G  N  Q  L  M
1081 GGAGCTGACAGCACACCAGAGCAGCGTGGGGAGCGTCCTGCAGGCTGGCAACCAGCTGAT 1140

    T  Q  G  T  L  S  E  E  E  E  F  E  I  Q  E  Q  M  T  L  L
1141 GACACAAGGGACTCTGTCAGAGGAGGAGGAGTTTGAGATCCAGGAACAGATGACCTTGCT 1200
```

Figure 9 Continued

```
          N  A  R  W  E  A  L  R  V  E  S  M  E  R  Q  S  R  L  H  D
1201  GAATGCAAGGTGGGAGGCGCTCCGGGTGGAGAGCATGGAGAGGCAGTCCCGGCTGCACGA  1260

          A  L  M  E  L  Q  K  K  Q  L  Q  Q  L  S  S  W  L  A  L  T
1261  CGCTCTGATGGAGCTGCAGAAGAAACAGCTGCAGCAGCTCTCAAGCTGGCTGGCCCTCAC  1320

          E  E  R  I  Q  K  M  E  S  P  P  L  G  D  D  L  P  S  L  Q
1321  AGAAGAGCGCATTCAGAAGATGGAGAGCCCTCCGCTGGGTGATGACCTGCCCTCCCTGCA  1380

          K  L  L  Q  E  H  K  S  L  Q  N  D  L  E  A  E  Q  V  K  V
1381  GAAGCTGCTTCAAGAACATAAAAGTTTGCAAAATGACCTTGAAGCTGAACAGGTGAAGGT  1440

          N  S  L  T  H  M  V  V  I  V  D  E  N  S  G  E  S  A  T  A
1441  AAATTCCTTAACTCACATGGTGGTGATTGTGGATGAAAACAGTGGGGAGAGTGCCACAGC  1500

          L  L  E  D  Q  L  Q  K  L  G  E  R  W  T  A  V  C  R  W  T
1501  TCTTCTGGAAGATCAGTTACAGAAACTGGGTGAGCGCTGGACAGCTGTATGCCGCTGGAC  1560

          E  E  R  W  N  R  L  Q  E  I  S  I  L  W  Q  E  L  L  E  E
1561  TGAAGAACGTTGGAACAGGTTGCAAGAAATCAGTATTCTGTGGCAGGAATTATTGGAAGA  1620

          Q  C  L  L  E  A  W  L  T  E  K  E  E  A  L  N  K  V  Q  T
1621  GCAGTGTCTGTTGGAGGCTTGGCTCACCGAAAAGGAAGAGGCTTTGAATAAAGTTCAAAC  1680

          S  N  F  K  D  Q  K  E  L  S  V  S  V  R  R  L  A  I  L  K
1681  CAGCAACTTTAAAGACCAGAAGGAACTAAGTGTCAGTGTCCGGCGTCTGGCTATATTGAA  1740

          E  D  M  E  M  K  R  Q  T  L  D  Q  L  S  E  I  G  Q  D  V
1741  GGAAGACATGGAAATGAAGAGGCAGACTCTGGATCAACTGAGTGAGATTGGCCAGGATGT  1800

          G  Q  L  L  S  N  P  K  A  S  K  K  M  N  S  D  S  E  E  L
1801  GGGCCAATTACTCAGTAATCCCAAGGCATCTAAGAAGATGAACAGTGACTCTGAGGAGCT  1860

          T  Q  R  W  D  S  L  V  Q  R  L  E  D  S  S  N  Q  V  T  Q
1861  AACACAGAGATGGGATTCTCTGGTTCAGAGACTCGAAGACTCTTCTAACCAGGTGACTCA  1920

          A  V  A  K  L  G  M  S  Q  I  P  Q  K  D  L  L  E  T  V  H
1921  GGCGGTAGCGAAGCTCGGCATGTCCCAGATTCCACAGAAGGACCTATTGGAGACCGTTCA  1980

          V  R  E  K  G  M  V  K  K  P  K  Q  E  L  P  P  P  L  G  P
1981  TGTGAGAGAAAAAGGGATGGTGAAGAAGCCCAAGCAGGAACTGCCTCCTCCGTTGGGCCC  2040

          K  K  R  Q  I  H  V  D  I  E  A  K  K  K  F  D  A  I  S  A
2041  AAAGAAGAGACAGATCCATGTGGATATTGAAGCTAAGAAAAAGTTTGATGCTATAAGTGC  2100

          E  L  L  N  W  I  L  K  W  K  T  A  I  Q  T  T  E  I  K  E
2101  AGAGCTGTTGAACTGGATTTTGAAATGGAAAACTGCCATTCAGACCACAGAGATAAAAGA  2160

          Y  M  K  M  Q  D  T  S  E  M  K  K  K  L  K  A  L  E  K  E
2161  GTATATGAAGATGCAAGACACTTCCGAAATGAAAAAGAAGTTGAAGGCATTAGAAAAAGA  2220

          Q  R  E  R  I  P  R  A  D  E  L  N  Q  T  G  Q  I  L  V  E
2221  ACAGAGAGAAAGAATCCCCAGAGCAGATGAATTAAACCAAACTGGACAAATCCTTGTGGA  2280

          Q  M  G  K  E  G  L  P  T  E  E  I  K  N  V  L  E  K  V  S
2281  GCAAATGGGAAAAGAAGGCCTTCCTACTGAAGAAATAAAAAATGTTCTGGAGAAGGTTTC  2340

          S  E  W  K  N  V  S  Q  H  L  E  D  L  E  R  K  I  Q  L  Q
2341  ATCAGAATGGAAGAATGTATCTCAACATTTGGAAGATCTAGAAAGAAAGATTCAGCTACA  2400

          E  D  I  N  A  Y  F  K  Q  L  D  E  L  E  K  V  I  K  T  K
2401  GGAAGATATAAATGCTTATTTCAAGCAGCTTGATGAGCTTGAAAAGGTCATCAAGACAAA  2460
```

Figure 9 Continued

```
          E  E  W  V  K  H  T  S  I  S  E  S  S  R  Q  S  L  P  S  L
2461  GGAGGAGTGGGTAAAACACACTTCCATTTCTGAATCTTCCCGGCAGTCCTTGCCAAGCTT  2520

          K  D  S  C  Q  R  E  L  T  N  L  L  G  L  H  P  K  I  E  M
2521  GAAGGATTCCTGTCAGCGGGAATTGACAAATCTTCTTGGCCTTCACCCCAAAATTGAAAT  2580

          A  R  A  S  C  S  A  L  M  S  Q  P  S  A  P  D  F  V  Q  R
2581  GGCTCGTGCAAGCTGCTCGGCCCTGATGTCTCAGCCTTCTGCCCCAGATTTTGTCCAGCG  2640

          G  F  D  S  F  L  G  R  Y  Q  A  V  Q  E  A  V  E  D  R  Q
2641  GGGCTTCGATAGCTTTCTGGGCCGCTACCAAGCTGTACAAGAGGCTGTAGAGGATCGTCA  2700

          Q  H  L  E  N  E  L  K  G  Q  P  G  H  A  Y  L  E  T  L  K
2701  ACAACATCTAGAGAATGAACTGAAGGGCCAACCTGGACATGCATATCTGGAAACATTGAA  2760

          T  L  K  D  V  L  N  D  S  E  N  K  A  Q  V  S  L  N  V  L
2761  AACACTGAAAGATGTGCTAAATGATTCAGAAAATAAGGCCCAGGTGTCTCTGAATGTCCT  2820

          N  D  L  A  K  V  E  K  A  L  Q  E  K  K  T  L  D  E  I  L
2821  TAATGATCTTGCCAAGGTGGAGAAGGCCCTGCAAGAAAAAAAGACCCTTGATGAAATCCT  2880

          E  N  Q  K  P  A  L  H  K  L  A  E  E  T  K  A  L  E  K  N
2881  TGAGAATCAGAAACCTGCATTACATAAACTTGCAGAAGAAACAAAGGCTCTGGAGAAAAA  2940

          V  H  P  D  V  E  K  L  Y  K  Q  E  F  D  D  V  Q  G  K  W
2941  TGTTCATCCTGATGTAGAAAAATTATATAAGCAAGAATTTGATGATGTGCAAGGAAAGTG  3000

          N  K  L  K  V  L  V  S  K  D  L  H  L  L  E  E  I  A  L  T
3001  GAACAAGCTAAAGGTCTTGGTTTCCAAAGATCTACATTTGCTTGAGGAAATTGCTCTCAC  3060

          L  R  A  F  E  A  D  S  T  V  I  E  K  W  M  D  G  V  K  D
3061  ACTCAGAGCTTTTGAGGCCGATTCAACAGTCATTGAGAAGTGGATGGATGGCGTGAAAGA  3120

          F  L  M  K  Q  Q  A  A  Q  G  D  D  A  G  L  Q  R  Q  L  D
3121  CTTCTTAATGAAACAGCAGGCTGCCCAAGGAGACGACGCAGGTCTACAGAGGCAGTTAGA  3180

          Q  C  S  A  F  V  N  E  I  E  T  I  E  S  S  L  K  N  M  K
3181  CCAGTGCTCTGCATTTGTTAATGAAATAGAAACAATTGAATCATCTCTGAAAAACATGAA  3240

          E  I  E  T  N  L  R  S  G  P  V  A  G  I  K  T  W  V  Q  T
3241  GGAAATAGAGACTAATCTTCGAAGTGGTCCAGTTGCTGGAATAAAAACTTGGGTGCAGAC  3300

          R  L  G  D  Y  Q  T  Q  L  E  K  L  S  K  E  I  A  T  Q  K
3301  AAGACTAGGTGACTACCAAACTCAACTGGAGAAACTTAGCAAGGAGATCGCTACTCAAAA  3360

          S  R  L  S  E  S  Q  E  K  A  A  N  L  K  K  D  L  A  E  M
3361  AAGTAGGTTGTCTGAAAGTCAAGAAAAAGCTGCGAACCTGAAGAAAGACTTGGCAGAGAT  3420

          Q  E  W  M  T  Q  A  E  E  E  Y  L  E  R  D  F  E  Y  K  S
3421  GCAGGAATGGATGACCCAGGCCGAGGAAGAATATTTGGAGCGGGATTTTGAGTACAAGTC  3480

          P  E  E  L  E  S  A  V  E  E  M  K  R  A  K  E  D  V  L  Q
3481  ACCAGAAGAGCTTGAGAGTGCTGTGGAAGAGATGAAGAGGGCAAAAGAGGATGTGTTGCA  3540

          K  E  V  R  V  K  I  L  K  D  N  I  K  L  L  A  A  K  V  P
3541  GAAGGAGGTGAGAGTGAAGATTCTCAAGGACAACATCAAGTTATTAGCTGCCAAGGTGCC  3600

          S  G  G  Q  E  L  T  S  E  L  N  V  V  L  E  N  Y  Q  L  L
3601  CTCTGGTGGCCAGGAGTTGACGTCTGAGCTGAATGTTGTGCTGGAGAATTACCAACTTCT  3660

          C  N  R  I  R  G  K  C  H  T  L  E  E  V  W  S  C  W  I  E
3661  TTGTAATAGAATTCGAGGAAAGTGCCACACGCTAgagGAGGTCTGGTCTTGTTGGATTGA  3720
```

Figure 9 Continued

```
          L  L  H  Y  L  D  L  E  T  T  W  L  N  T  L  E  E  R  M  K
3721  ACTGCTTCACTATTTGGATCTTGAAACTACCTGGTTAAACACTTTGGAAGAGCGGATGAA  3780

          S  T  E  V  L  P  E  K  T  D  A  V  N  E  A  L  E  S  L  E
3781  GAGCACAGAGGTCCTGCCTGAGAAGACGGATGCTGTCAACGAAGCCCTGGAGTCTCTGGA  3840

          S  V  L  R  H  P  A  D  N  R  T  Q  I  R  E  L  G  Q  T  L
3841  ATCTGTTCTGCGCCACCCGGCAGATAATCGCACCCAGATTCGAGAGCTTGGCCAGACTCT  3900

          I  D  G  G  I  L  D  D  I  I  S  E  K  L  E  A  F  N  S  R
3901  GATTGATGGGGGGATCCTGGATGATATAATCAGTGAGAAACTGGAGGCTTTCAACAGCCG  3960

          Y  E  D  L  S  H  L  A  E  S  K  Q  I  S  L  E  K  Q  L  Q
3961  ATATGAAGATCTAAGTCACCTGGCAGAGAGCAAGCAGATTTCTTTGGAAAAGCAACTCCA  4020

          V  L  R  E  T  D  Q  M  L  Q  V  L  Q  E  S  L  G  E  L  D
4021  GGTGCTGCGGGAAACTGACCAGATGCTTCAAGTCTTGCAAGAGAGCTTGGGGGAGCTGGA  4080

          K  Q  L  T  T  Y  L  T  D  R  I  D  A  F  Q  V  P  Q  E  A
4081  CAAACAGCTCACCACATACCTGACTGACAGGATAGATGCTTTCCAAGTTCCACAGGAAGC  4140

          Q  K  I  Q  A  E  I  S  A  H  E  L  T  L  E  E  L  R  R  N
4141  TCAGAAAATCCAAGCAGAGATCTCAGCCCATGAGCTAACCCTAGAGGAGTTGAGAAGAAA  4200

          M  R  S  Q  P  L  T  S  P  E  S  R  T  A  R  G  G  S  Q  M
4201  TATGCGTTCTCAGCCCCTGACCTCCCCAGAGAGTAGGACTGCCAGAGGAGGAAGTCAGAT  4260

          D  V  L  Q  R  K  L  R  E  V  S  T  K  F  Q  L  F  Q  K  P
4261  GGATGTGCTACAGAGGAAACTCCGAGAGGTGTCCACAAAGTTCCAGCTTTTCCAGAAGCC  4320

          A  N  F  E  Q  R  M  L  D  C  K  R  V  L  D  G  V  K  A  E
4321  AGCTAACTTCGAGCAGCGCATGCTGGACTGCAAGCGTGTGCTGGATGGCGTGAAAGCAGA  4380

          L  H  V  L  D  V  K  D  V  D  P  D  V  I  Q  T  H  L  D  K
4381  ACTTCACGTTCTGGATGTGAAGGACGTAGACCCTGACGTCATACAGACGCACCTGGACAA  4440

          C  M  K  L  Y  K  T  L  S  E  V  K  L  E  V  E  T  V  I  K
4441  GTGTATGAAACTGTATAAAACTTTGAGTGAAGTCAAACTTGAAGTGGAAACTGTGATTAA  4500

          T  G  R  H  I  V  Q  K  Q  Q  T  D  N  P  K  G  M  D  E  Q
4501  AACAGGAAGACATATTGTCCAGAAACAGCAAACGGACAACCCAAAAGGGATGGATGAGCA  4560

          L  T  S  L  K  V  L  Y  N  D  L  G  A  Q  V  T  E  G  K  Q
4561  GCTGACTTCCCTGAAGGTTCTTTACAATGACCTGGGCGCACAGGTGACAGAAGGAAAACA  4620

          D  L  E  R  A  S  Q  L  A  R  K  M  K  K  E  A  A  S  L  S
4621  GGATCTGGAAAGAGCATCACAGTTGGCCCGGAAAATGAAGAAAGAGGCTGCTTCTCTCTC  4680

          E  W  L  S  A  T  E  T  E  L  V  Q  K  S  T  S  E  G  L  L
4681  TGAATGGCTTTCTGCTACTGAAACTGAATTGGTACAGAAGTCCACTTCAGAAGGTCTGCT  4740

          G  D  L  D  T  E  I  S  W  A  K  N  V  L  K  D  L  E  K  R
4741  TGGTGACTTGGATACAGAAATTTCCTGGGCTAAAAATGTTCTGAAGGATCTGGAAAAGAG  4800

          K  A  D  L  N  T  I  T  E  S  S  A  A  L  Q  N  L  I  E  G
4801  AAAAGCTGATTTAAATACCATCACAGAGAGTAGTGCTGCCCTGCAAAACTTGATTGAGGG  4860

          S  E  P  I  L  E  E  R  L  C  V  L  N  A  G  W  S  R  V  R
4861  CAGTGAGCCTATTTTAGAAGAGAGGCTCTGCGTCCTTAACGCTGGGTGGAGCCGAGTTCG  4920

          T  W  T  E  D  W  C  N  T  L  M  N  H  Q  N  Q  L  E  I  F
4921  TACCTGGACTGAAGATTGGTGCAATACCTTGATGAACCATCAGAACCAGCTAGAAATATT  4980
```

Figure 9 Continued

```
          D  G  N  V  A  H  I  S  T  W  L  Y  Q  A  E  A  L  L  D  E
4981  TGATGGGAACGTGGCTCACATAAGTACCTGGCTTTATCAAGCTGAAGCTCTATTGGATGA  5040

          I  E  K  K  P  T  S  K  Q  E  E  I  V  K  R  L  V  S  E  L
5041  AATTGAAAAGAAACCAACAAGTAAACAGGAAGAAATTGTGAAGCGTTTAGTATCTGAGCT  5100

          D  D  A  N  L  Q  V  E  N  V  R  D  Q  A  L  I  L  M  N  A
5101  GGATGATGCCAACCTCCAGGTTGAAAATGTCCGCGATCAAGCCCTTATTTTGATGAATGC  5160

          R  G  S  S  S  R  E  L  V  E  P  K  L  A  E  L  N  R  N  F
5161  CCGTGGAAGCTCAAGCAGGGAGCTTGTAGAACCAAAGTTAGCTGAGCTGAATAGGAACTT  5220

          E  K  V  S  Q  H  I  K  S  A  K  L  L  I  A  Q  E  P  L  Y
5221  TGAAAAGGTGTCTCAACATATCAAAAGTGCCAAATTGCTAATTGCTCAGGAACCATTATA  5280

          Q  C  L  V  T  T  E  T  F  E  T  G  V  P  F  S  D  L  E  K
5281  CCAATGTTTGGTCACCACTGAAACATTTGAAACTGGTGTGCCTTTCTCTGACTTGGAAAA  5340

          L  E  N  D  I  E  N  M  L  K  F  V  E  K  H  L  E  S  S  D
5341  ATTAGAAAATGACATAGAAAATATGTTAAAATTTGTGGAAAAACACTTGGAATCCAGTGA  5400

          E  D  E  K  M  D  E  E  S  A  Q  I  E  E  V  L  Q  R  G  E
5401  TGAAGATGAAAAGATGGATGAGGAGAGTGCCCAGATTGAGGAAGTTCTACAAAGAGGAGA  5460

          E  M  L  H  Q  P  M  E  D  N  K  K  E  K  I  R  L  Q  L  L
5461  AGAAATGTTACATCAACCTATGGAAGATAATAAAAAAGAAAAGATCCGTTTGCAATTATT  5520

          L  L  H  T  R  Y  N  K  I  K  A  I  P  I  Q  Q  R  K  M  G
5521  ACTTTTGCATACTAGATACAACAAAATTAAGGCAATCCCTATTCAACAGAGGAAAATGGG  5580

          Q  L  A  S  G  I  R  S  S  L  L  P  T  D  Y  L  V  E  I  N
5581  TCAACTTGCTTCTGGAATTAGATCATCACTTCTTCCTACAGATTATCTGGTTGAAATTAA  5640

          K  I  L  L  C  M  D  D  V  E  L  S  L  N  V  P  E  L  N  T
5641  CAAAATTTTACTTTGCATGGATGATGTTGAATTATCGCTTAATGTTCCAGAGCTCAACAC  5700

          A  I  Y  E  D  F  S  F  Q  E  D  S  L  K  N  I  K  D  Q  L
5701  TGCTATTTACGAAGACTTCTCTTTTCAGGAAGACTCTCTGAAGAATATCAAAGACCAACT  5760

          D  K  L  G  E  Q  I  A  V  I  H  E  K  Q  P  D  V  I  L  E
5761  GGACAAACTTGGAGAGCAGATTGCAGTCATTCATGAAAAACAGCCAGATGTCATCCTTGA  5820

          A  S  G  P  E  A  I  Q  I  R  D  T  L  T  Q  L  N  A  K  W
5821  AGCCTCTGGACCTGAAGCCATTCAGATCAGAGATACACTTACTCAGCTGAATGCAAAATG  5880

          D  R  I  N  R  M  Y  S  D  R  K  G  C  F  D  R  A  M  E  E
5881  GGACAGAATTAATAGAATGTACAGTGATCGGAAAGGTTGTTTTGACAGGGCAATGGAAGA  5940

          W  R  Q  F  H  C  D  L  N  D  L  T  Q  W  I  T  E  A  E  E
5941  ATGGAGACAGTTCCATTGTGACCTTAATGACCTCACACAGTGGATAACAGAGGCTGAAGA  6000

          L  L  V  D  T  C  A  P  G  G  S  L  D  L  E  K  A  R  I  H
6001  ATTACTGGTTGATACCTGTGCTCCAGGTGGCAGCCTGGACTTAGAGAAAGCCAGGATACA  6060

          Q  Q  E  L  E  V  G  I  S  S  H  Q  P  S  F  A  A  L  N  R
6061  TCAGCAGGAACTTGAGGTGGGCATCAGCAGCCACCAGCCCAGTTTTGCAGCACTAAACCG  6120

          T  G  D  G  I  V  Q  K  L  S  Q  A  D  G  S  F  L  K  E  K
6121  AACTGGGGATGGGATTGTGCAGAAACTCTCCCAGGCAGATGGAAGCTTCTTGAAAGAAAA  6180

          L  A  G  L  N  Q  R  W  D  A  I  V  A  E  V  K  D  R  Q  P
6181  ACTGGCAGGTTTAAACCAACGCTGGGATGCAATTGTTGCAGAAGTGAAGGATAGGCAGCC  6240
```

Figure 9 Continued

```
          R  L  K  G  E  S  K  Q  V  M  K  Y  R  H  Q  L  D  E  I  I
6241  AAGGCTAAAAGGAGAAAGTAAGCAGGTGATGAAGTACAGGCATCAGCTAGATGAGATTAT  6300

          C  W  L  T  K  A  E  H  A  M  Q  K  R  S  T  T  E  L  G  E
6301  CTGTTGGTTAACAAAGGCTGAGCATGCTATGCAAAAGAGATCAACCACCGAATTGGGAGA  6360

          N  L  Q  E  L  R  D  L  T  Q  E  M  E  V  H  A  E  K  L  K
6361  AAACCTGCAAGAATTAAGAGACTTAACTCAAGAAATGGAAGTACATGCTGAAAAACTCAA  6420

          W  L  N  R  T  E  L  E  M  L  S  D  K  S  L  S  L  P  E  R
6421  ATGGCTGAATAGAACTGAATTGGAGATGCTTTCAGATAAAAGTCTGAGTTTACCTGAAAG  6480

          D  K  I  S  E  S  L  R  T  V  N  M  T  W  N  K  I  C  R  E
6481  GGATAAAATTTCAGAAAGCTTAAGGACTGTAAATATGACATGGAATAAGATTTGCAGAGA  6540

          V  P  T  T  L  K  E  C  I  Q  E  P  S  S  V  S  Q  T  R  I
6541  GGTGCCTACCACCCTGAAGGAATGCATCCAGGAGCCCAGTTCTGTTTCACAGACAAGGAT  6600

          A  A  H  P  N  V  Q  K  V  V  L  V  S  S  A  S  D  I  P  V
6601  TGCTGCTCATCCTAATGTCCAAAAGGTGGTGCTAGTATCATCTGCGTCAGATATTCCTGT  6660

          Q  S  H  R  T  S  E  I  S  I  P  A  D  L  D  K  T  I  T  E
6661  TCAGTCTCATCGTACTTCGGAAATTTCAATTCCTGCTGATCTTGATAAAACTATAACAGA  6720

          L  A  D  W  L  V  L  I  D  Q  M  L  K  S  N  I  V  T  V  G
6721  ACTAGCCGACTGGCTGGTATTAATCGACCAGATGCTGAAGTCCAACATTGTCACTGTTGG  6780

          D  V  E  E  I  N  K  T  V  S  R  M  K  I  T  K  A  D  L  E
6781  GGATGTAGAAGAGATCAATAAGACCGTTTCCCGAATGAAAATTACAAAGGCTGACTTAGA  6840

          Q  R  H  P  Q  L  D  Y  V  F  T  L  A  Q  N  L  K  N  K  A
6841  ACAGCGCCATCCTCAGCTGGATTATGTTTTTACATTGGCACAGAATTTGAAAAATAAAGC  6900

          S  S  S  D  M  R  T  A  I  T  E  K  L  E  R  V  K  N  Q  W
6901  TTCCAGTTCAGATATGAGAACAGCAATTACAGAAAAATTGGAAAGGGTCAAGAACCAGTG  6960

          D  G  T  Q  H  G  V  E  L  R  Q  Q  Q  L  E  D  M  I  I  D
6961  GGATGGCACCCAGCATGGCGTTGAGCTAAGACAGCAGCAGCTTGAGGACATGATTATTGA  7020

          S  L  Q  W  D  D  H  R  E  E  T  E  E  L  M  R  K  Y  E  A
7021  CAGTCTTCAGTGGGATGACCATAGGGAGGAGACTGAAGAACTGATGAGAAAATATGAGGC  7080

          R  L  Y  I  L  Q  Q  A  R  R  D  P  L  T  K  Q  I  S  D  N
7081  TCGACTCTATATTCTTCAGCAAGCCCGACGGGATCCACTCACCAAACAAATTTCTGATAA  7140

          Q  I  L  L  Q  E  L  G  P  G  D  G  I  V  M  A  F  D  N  V
7141  CCAAATACTGCTTCAAGAACTGGGTCCTGGAGATGGTATCGTCATGGCGTTCGATAACGT  7200

          L  Q  K  L  L  E  E  Y  G  S  D  D  T  R  N  V  K  E  T  T
7201  CCTGCAGAAACTCCTGGAGGAATATGGGAGTGATGACACAAGGAATGTGAAAGAAACCAC  7260

          E  Y  L  K  T  S  W  I  N  L  K  Q  S  I  A  D  R  Q  N  A
7261  AGAGTACTTAAAAACATCATGGATCAATCTCAAACAAAGTATTGCTGACAGACAGAACGC  7320

          L  E  A  E  W  R  T  V  Q  A  S  R  R  D  L  E  N  F  L  K
7321  CTTGGAGGCTGAGTGGAGGACGGTGCAGGCCTCTCGCAGAGATCTGGAAAACTTCCTGAA  7380

          W  I  Q  E  A  E  T  T  V  N  V  L  V  D  A  S  H  R  E  N
7381  GTGGATCCAAGAAGCAGAGACCACAGTGAATGTGCTTGTGGATGCCTCTCATCGGGAGAA  7440

          A  L  Q  D  S  I  L  A  R  E  L  K  Q  Q  M  Q  D  I  Q  A
7441  TGCTCTTCAGGATAGTATCTTGGCCAGGGAACTCAAACAGCAGATGCAGGACATCCAGGC  7500
```

Figure 9 Continued


```
      E  I  D  A  H  N  D  I  F  K  S  I  D  G  N  R  Q  K  M  V
7501  AGAAATTGATGCCCACAATGACATATTTAAAAGCATTGACGGAAACAGGCAGAAGATGGT  7560

      K  A  L  G  N  S  E  E  A  T  M  L  Q  H  R  L  D  D  M  N
7561  AAAAGCTTTGGGAAATTCTGAAGAGGCTACTATGCTTCAACATCGACTGGATGATATGAA  7620

      Q  R  W  N  D  L  K  A  K  S  A  S  I  R  A  H  L  E  A  S
7621  CCAAAGATGGAATGACTTAAAAGCAAAATCTGCTAGCATCAGGGCCCATTTGGAGGCCAG  7680

      A  E  K  W  N  R  L  L  M  S  L  E  E  L  I  K  W  L  N  M
7681  CGCTGAGAAGTGGAACAGGTTGCTGATGTCCTTAGAAGAACTGATCAAATGGCTGAATAT  7740

      K  D  E  E  L  K  K  Q  M  P  I  G  G  D  V  P  A  L  Q  L
7741  GAAAGATGAAGAGCTTAAGAAACAAATGCCTATTGGAGGAGATGTTCCAGCCTTACAGCT  7800

      Q  Y  D  H  C  K  A  L  R  R  E  L  K  E  K  E  Y  S  V  L
7801  CCAGTATGACCATTGTAAGGCCCTGAGACGGGAGTTAAAGGAGAAAGAATATTCTGTCCT  7860

      N  A  V  D  Q  A  R  V  F  L  A  D  Q  P  I  E  A  P  E  E
7861  GAATGCTGTCGACCAGGCCCGAGTTTTCTTGGCTGATCAGCCAATTGAGGCCCCTGAAGA  7920

      P  R  R  N  L  Q  S  K  T  E  L  T  P  E  E  R  A  Q  K  I
7921  GCCAAGAAGAAACCTACAATCAAAAACAGAATTAACTCCTGAGGAGAGAGCCCAAAAGAT  7980

      A  K  A  M  R  K  Q  S  S  E  V  K  E  K  W  E  S  L  N  A
7981  TGCCAAAGCCATGCGCAAACAGTCTTCTGAAGTCAAAGAAAAATGGGAAAGTCTAAATGC  8040

      V  T  S  N  W  Q  K  Q  V  D  K  A  L  E  K  L  R  D  L  Q
8041  TGTAACTAGCAATTGGCAAAAGCAAGTGGACAAGGCATTGGAGAAACTCAGAGACCTGCA  8100

      G  A  M  D  D  L  D  A  D  M  K  E  A  E  S  V  R  N  G  W
8101  GGGAGCTATGGATGACCTGGACGCTGACATGAAGGAGGCAGAGTCCGTGCGGAATGGCTG  8160

      K  P  V  G  D  L  L  I  D  S  L  Q  D  H  I  E  K  I  M  A
8161  GAAGCCCGTGGGAGACTTACTCATTGACTCGCTGCAGGATCACATTGAAAAAATCATGGC  8220

      F  R  E  E  I  A  P  I  N  F  K  V  K  T  V  N  D  L  S  S
8221  ATTTAGAGAAGAAATTGCACCAATCAACTTTAAAGTTAAAACGGTGAATGATTTATCCAG  8280

      Q  L  S  P  L  D  D  L  H  P  S  L  K  M  S  R  Q  L  D  D  L
8281  TCAGCTGTCTCCACTTGACCTGCATCCCTCTCTAAAGATGTCTCGCCAGCTAGATGACCT  8340

      N  M  R  W  K  L  L  Q  V  S  V  D  D  R  L  K  Q  L  Q  E
8341  TAATATGCGATGGAAACTTTTACAGGTTTCTGTGGATGATCGCCTTAAACAGCTTCAGGA  8400

      A  H  R  D  F  G  P  S  S  Q  H  F  L  S  T  S  V  Q  L  P
8401  AGCCCACAGAGATTTTGGACCATCCTCTCAGCATTTTCTCTCTACGTCAGTCCAGCTGCC  8460

      W  Q  R  S  I  S  H  N  K  V  P  Y  Y  I  N  H  Q  T  Q  T
8461  GTGGCAAAGATCCATTTCACATAATAAAGTGCCCTATTACATCAACCATCAAACACAGAC  8520

      T  C  W  D  H  P  K  M  T  E  L  F  Q  S  L  A  D  L  N  N
8521  CACCTGTTGGGACCATCCTAAAATGACCGAACTCTTTCAATCCCTTGCTGACCTGAATAA  8580

      V  R  F  S  A  Y  R  T  A  I  K  I  R  R  L  Q  K  A  L  C
8581  TGTACGTTTTTCTGCCTACCGTACAGCAATCAAAATCCGAAGACTACAAAAAGCACTATG  8640

      L  D  L  L  E  L  S  T  T  N  E  I  F  K  Q  H  K  L  N  Q
8641  TTTGGATCTCTTAGAGTTGAGTACAACAAATGAAATTTTCAAACAGCACAAGTTGAACCA  8700

      N  D  Q  L  L  S  V  P  D  V  I  N  C  L  T  T  T  Y  D  G
8701  AAATGACCAGCTCCTCAGTGTTCCAGATGTCATCAACTGTCTGACAACAACTTATGATGG  8760
```

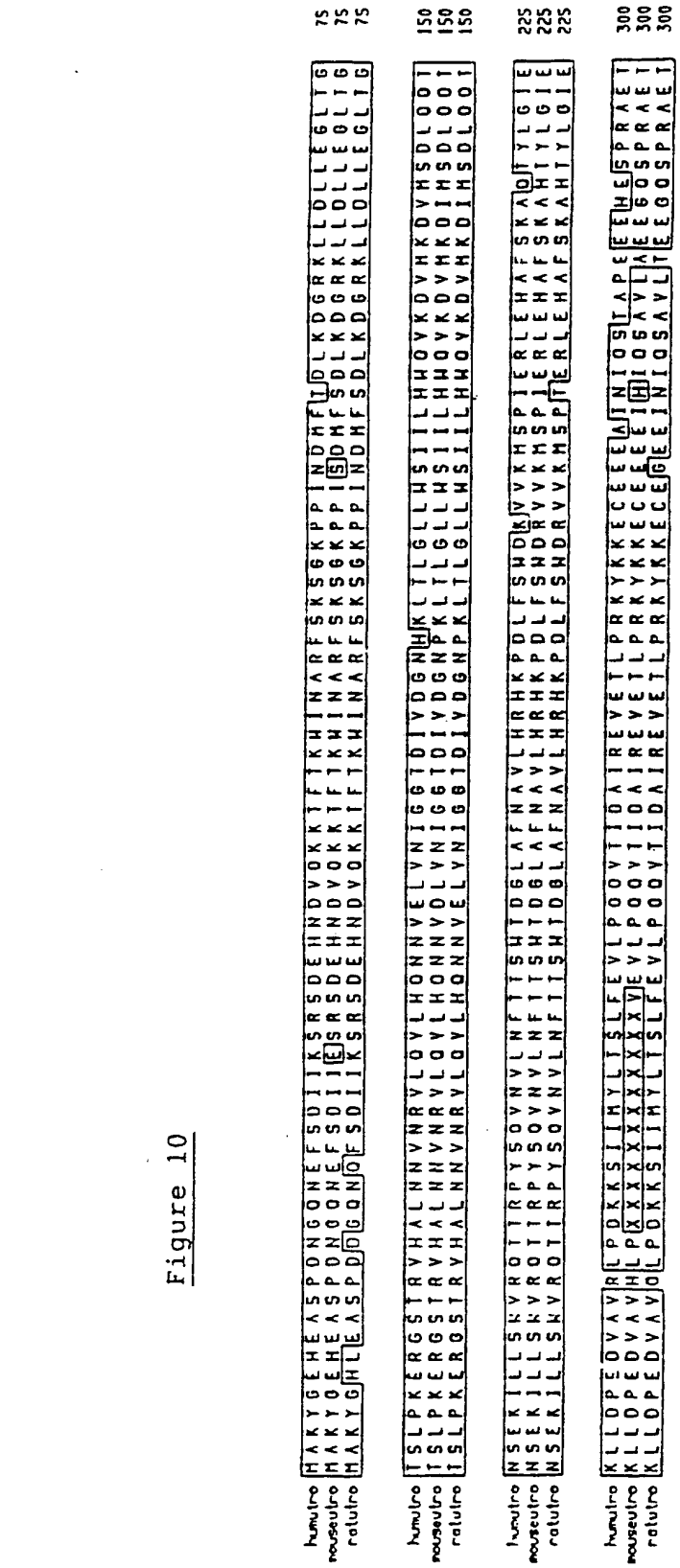
Figure 9 Continued

```
          L  E  Q  M  H  K  D  L  V  N  V  P  L  C  V  D  M  C  L  N
8761  ACTTGAGCAAATGCATAAGGACCTGGTCAACGTTCCACTCTGTGTTGATATGTGTCTCAA  8820

          W  L  L  N  V  Y  D  T  G  R  T  G  K  I  R  V  Q  S  L  K
8821  TTGGTTGCTCAATGTCTATGACACGGGTCGAACTGGAAAAATTAGAGTGCAGAGTCTGAA  8880

          I  G  L  M  S  L  S  K  G  L  L  E  E  K  Y  R  Y  L  F  K
8881  GATTGGATTAATGTCTCTCTCCAAAGGTCTCTTGGAAGAAAAATACAGATATCTCTTTAA  8940

          E  V  A  G  P  T  E  M  C  D  Q  R  Q  L  G  L  L  L  H  D
8941  GGAAGTTGCGGGGCCGACAGAAATGTGTGACCAGAGGCAGCTGGGCCTGTTACTTCATGA  9000

          A  I  Q  I  P  R  Q  L  G  E  V  A  A  F  G  G  S  N  I  E
9001  TGCCATCCAGATCCCCCGGCAGCTAGGTGAAGTAGCAGCTTTTGGAGGCAGTAATATTGA  9060

          P  S  V  R  S  C  F  Q  Q  N  N  N  K  P  E  I  S  V  K  E
9061  GCCTAGTGTTCGCAGCTGCTTCCAACAGAATAACAATAAACCAGAAATAAGTGTGAAAGA  9120

          F  I  D  W  M  H  L  E  P  Q  S  M  V  W  L  P  V  L  H  R
9121  GTTTATAGATTGGATGCATTTGGAACCACAGTCCATGGTTTGGCTCCCAGTTTTACATCG  9180

          V  A  A  A  E  T  A  K  H  Q  A  K  C  N  I  C  K  E  C  P
9181  AGTGGCAGCAGCGGAGACTGCAAAACATCAGGCCAAATGCAACATCTGTAAAGAATGTCC  9240

          I  V  G  F  R  Y  R  S  L  K  H  F  N  Y  D  V  C  Q  S  C
9241  AATTGTCGGGTTCAGGTATAGAAGCCTTAAGCATTTTAACTATGATGTCTGCCAGAGTTG  9300

          F  F  S  G  R  T  A  K  G  H  K  L  H  Y  P  M  V  E  Y  C
9301  TTTCTTTTCGGGTCGAACAGCAAAAGGTCACAAATTACATTACCCAATGGTGGAATATTG  9360

          I  P  T  T  S  G  E  D  V  R  D  F  T  K  V  L  K  N  K  F
9361  TATACCTACAACATCTGGGGAAGATGTACGAGACTTCACAAAGGTACTTAAGAACAAGTT  9420

          R  S  K  K  Y  F  A  K  H  P  R  L  G  Y  L  P  V  Q  T  V
9421  CAGGTCGAAGAAGTACTTTGCCAAACACCCTCGACTTGGTTACCTGCCTGTCCAGACAGT  9480

          L  E  G  D  N  L  E  T  P  I  T  L  I  S  M  W  P  E  H  Y
9481  TCTTGAAGGTGACAACTTAGAGACTCCTATCACACTCATCAGTATGTGGCCAGAGCACTA  9540

          D  P  S  Q  S  P  Q  L  F  H  D  D  T  H  S  R  I  E  Q  Y
9541  TGACCCCTCACAATCTCCTCAACTGTTTCATGATGACACCCATTCAAGAATAGAACAATA  9600

          A  T  R  L  A  Q  M  E  R  T  N  G  S  F  L  T  D  S  S  S
9601  TGCCACACGACTGGCCCAGATGGAAAGGACTAATGGGTCTTTTCTCACTGATAGCAGCTC  9660

          T  T  G  S  V  E  D  E  H  A  L  I  Q  Q  Y  C  Q  T  L  G
9661  CACCACAGGAAGTGTGGAAGACGAGCACGCCCTCATCCAGCAGTATTGCCAAACACTCGG  9720

          G  E  S  P  V  S  Q  P  Q  S  P  A  Q  I  L  K  S  V  E  R
9721  AGGAGAGTCCCCAGTGAGCCAGCCGCAGAGCCCAGCTCAGATCCTGAAGTCAGTAGAGAG  9780

          E  E  R  G  E  L  E  R  I  I  A  D  L  E  E  E  Q  R  N  L
9781  GGAAGAACGTGGAGAACTGGAGAGGATCATTGCTGACCTGGAGGAAGAACAAAGAAATCT  9840

          Q  V  E  Y  E  Q  L  K  D  Q  H  L  R  R  G  L  P  V  G  S
9841  ACAGGTGGAGTATGAGCAGCTGAAGGACCAGCACCTCCGAAGGGGGCTCCCTGTCGGTTC  9900

          P  P  E  S  I  I  S  P  H  H  T  S  E  D  S  E  L  I  A  E
9901  ACCGCCAGAGTCGATTATATCTCCCCATCACACGTCTGAGGATTCAGAACTTATAGCAGA  9960

          A  K  L  L  R  Q  H  K  G  R  L  E  A  R  M  Q  I  L  E  D
9961  AGCAAAACTCCTCAGGCAGCACAAAGGTCGGCTGGAGGCTAGGATGCAGATTTTAGAAGA  10020
```

Figure 9 Continued

```
          H  N  K  Q  L  E  S  Q  L  H  R  L  R  Q  L  L  E  Q  P  E
10021  TCACAATAAACAGCTGGAGTCTCAGCTCCACCGCCTCCGACAGCTGCTGGAGCAGCCTGA  10080

          S  D  S  R  I  N  G  V  S  P  W  A  S  P  Q  H  S  A  L  S
10081  ATCTGATTCCCGAATCAATGGTGTTTCCCCATGGGCTTCTCCTCAGCATTCTGCACTGAG  10140

          Y  S  L  D  P  D  A  S  G  P  Q  F  H  Q  A  A  G  E  D  L
10141  CTACTCGCTTGATCCAGATGCCTCCGGCCCACAGTTCCACCAGGCAGCGGGAGAGGACCT  10200

          L  A  P  P  H  D  T  S  T  D  L  T  E  V  M  E  Q  I  H  S
10201  GCTGGCCCCACCGCACGACACCAGCACGGATCTCACGGAGGTCATGGAGCAGATTCACAG  10260

          T  F  P  S  C  C  P  N  V  P  S  R  P  Q  A  M  *
10261  CACGTTTCCATCTTGCTGCCCAAATGTTCCCAGCAGGCCACAGGCAATGTAATCACTAGT  10320
```

Figure 10



A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12N15/79 C07K14/47 A61K31/70 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 92, April 1995, pages 3697-3701, XP002031008 BLAKE, D.J., ET AL . : "G-UTROPHIN, THE AUTOSOMAL HOMOLOGUE OF DYSTROPHIN Dp116, IS EXPRESSED IN SENSORY GANGLIA AND BRAIN" see the whole document & EMBL SEQUENCE DATA LIBRARY, 1 May 1995, HEIDELBERG, GERMANY, BLAKE, D.J., ET AL . : "G-UTROPHIN, THE AUTOSOMAL HOMOLOGUE OF DYSTROPHIN Dp 116, IS EXPRESSED IN SENSORY GANGLIA" ACCESSION No. x83506 --- -/--	3,7,8, 11,17,24

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 May 1997	27.05.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Holtorf, S

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NATURE, vol. 339, 4 May 1989, pages 55-58, XP002031004 LOVE, D.R., ET AL . : "AN AUTOSOMAL TRANSCRIPT IN SKELETAL MUSCLE WITH HOMOLOGY TO DYSTROPHIN" see the whole document & EMBL SEQUENCE DATA LIBRARY, 6 November 1992, HEIDELBERG, GERMANY, TINSLEY, J.M., ET AL .: "PRIMARY STRUCTURE OF DYSTROPHIN RELATED PROTEIN" ACCESSION No. X69086	3,6,8, 11,17
X	NATURE, vol. 360, 10 December 1992, pages 591-593, XP002031005 TINSLEY, J.M., ET AL . : "PRIMARY STRUCTURE OF DYSTROPHIN-RELATED PROTEIN" cited in the application see the whole document	7,8,11, 17
X	HUMAN MOLECULAR GENETICS, vol. 4, no. 8, August 1995, pages 1251-1258, XP002031006 PHELPS, S.F., ET AL . : "EXPRESSION OF FULL-LENGTH AND TRUNCATED DYSTROPHIN MINI-GENES IN TRANSGENIC mdx MICE" see the whole document	3,7-9, 11-21, 24,27,29
P,X	NATURE, vol. 384, 28 November 1996, pages 349-353, XP002031010 TINSLEY, J.M., ET AL . : "AMELIORATION OF THE DYSTROPHIC PHENOTYPE OF mdx MICE USING A TRUNCATED UTROPHIN TRANSGENE" see the whole document	1,2,4, 8-21,24, 25,27-29
A	NEUROMUSCULAR DISORDERS, vol. 3, no. 1, January 1993, pages 5-21, XP000673607 LOVE,D.R., ET AL . : "DYSTROPHIN AND DYSTROPHIN-RELATED PROTEINS: A REVIEW OF PROTEIN AND RNA STUDIES" see the whole document	1-30
A	CELL, vol. 53, 22 April 1988, pages 219-228, XP002031011 KÖNIG, M., ET AL . : "THE COMPLETE SEQUENCE OF DYSTROPHIN PREDICTS A ROD-SHAPED CYTOSKELETAL PROTEIN" see the whole document	1-30

-/--

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NEUROMUSCULAR DISORDERS, vol. 3, no. 5/6, 1993, pages 537-539, XP000578145 TINSLEY, J.M., ET AL . : "UTROPHIN: A POTENTIAL REPLACEMENT FOR DYSTROPHIN ?" see the whole document ---	1-30
A	FEBS LETTERS, vol. 358, 1995, pages 262-266, XP000486346 THI MAN, N., ET AL . : "FULL-LENGTH AND SHORT FORMS OF UTROPHIN, THE DYSTROPHIN-RELATED PROTEIN" see the whole document -----	1-30

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 96/03156

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 29,17,18,20-23,27,28
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although claims 29 and 17,18,20-23,27,28 as far as in vivo methods are concerned are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.